

STIC-Biotech/ChemLib

80355

From: Yu, Misook
Sent: Monday, November 18, 2002 8:26 AM
To: STIC-Biotech/ChemLib
Subject: RE: 09/191,497

09 919 497

Please disregard this request. I give you a wrong serial no. Thank you.

Examiner Misook Yu, Ph.D.
703-308-2454 (Phone)
Art Unit 1642
CM1-8E18 (Room)
CM1-8E12 (Mail Box)

Point of Contact:
Toby Port
Technical Info. Specialist
CM1 6A04
703-308-3534

-----Original Message-----

From: STIC-Biotech/ChemLib
Sent: Monday, November 18, 2002 8:03 AM
To: Yu, Misook
Subject: RE: 09/191,497

This Application Number does not have a valid CRF - please provide the necessary correction or another application number.

MAUDE

-----Original Message-----

From: Yu, Misook
Sent: Sunday, November 17, 2002 1:30 PM
To: STIC-Biotech/ChemLib
Subject: 09/191,497

Please search SEQ ID NO:14 and 32

Examiner Misook Yu, Ph.D.
703-308-2454 (Phone)
Art Unit 1642
CM1-8E18 (Room)
CM1-8E12 (Mail Box)

Searcher: _____
Phone: _____
Location: _____
Date Picked Up: 11/18
Date Completed: 11/27
Searcher Prep/Review: _____
Clerical: _____
Online time: _____

TYPE OF SEARCH:

NA Sequences: _____
AA Sequences: _____
Structures: _____
Bibliographic: _____
Litigation: _____
Full text: _____
Patent Family: _____
Other: _____

VENDOR/COST (where applic.)

STN: _____
DIALOG: _____
Questel/Orbit: _____
DRLink: _____
Lexis/Nexis: _____
Sequence Sys.: _____
WWW/Internet: _____
Other (specify): _____

THIS PAGE BLANK (USPTO)

**PALM INTRANET**Day : Monday
Date: 11/18/2002
Time: 07:19:25

Biotech Query for 09/191497

Title: **WEB-BASED ENGINEERING DESIGN AGENTS**Inventor: **RADCLIFFE, CLARK J.**Location: **9200/FILE REPOSITORY (FRANCONIA)**Location Date: **09/20/2001**Group Art Unit: **2177**Status: **150/PATENTED CASE**

Num	Date	Code	Contents Description
NO BIOTECH DATA			

Search for Biotech Info: Application# PCT / /

To go back use Back button on your browser toolbar.

Back to [PALM](#) | [ASSIGNMENT](#) | [OASIS](#) | [Home page](#)*auto 8:03*

THIS PAGE BLANK (USPTO)

STIC-Biotech/ChemLib

80355

From: Yu, Misook
Sent: Sunday, November 17, 2002 1:30 PM
To: STIC-Biotech/ChemLib
Subject: 09/191,497

mej

09 919 497

Please search SEQ ID NO:14 and 32

Examiner Misook Yu, Ph.D.
703-308-2454 (Phone)
Art Unit 1642
CM1-8E18 (Room)
CM1-8E12 (Mail Box)

TYPE OF SEARCH:

Searcher: _____
Phone: _____
Location: _____
Date Picked Up: _____
Date Completed: _____
Searcher Prep/Review: _____
Clerical: _____
Online time: _____

NA Sequences: _____
AA Sequences: _____
Structures: _____
Bibliographic: _____
Litigation: _____
Full text: _____
Patent Family: _____
Other: _____

VENDOR/COST (where applic.)

STN: _____
DIALOG: _____
Questel/Orbit: _____
DRLink: _____
Lexis/Nexis: _____
Sequence Sys.: _____
WWW/Internet: _____
Other (specify): _____

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 26, 2002, 19:44:35 ; Search time 6670.76 Seconds
(without alignments)
16827.091 Million cell updates/sec

Title: US-09-919-497-14

Perfect score: 3857

Sequence: 1 ggaagagcttcacatgagctc.....agaatacaacctaagtca 3857

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 segs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: gb_ba:*
2: gb_bt:*
3: gb_in:*
4: gb_ov:*
5: gb_ov:*
6: gb_ov:*
7: gb_ov:*
8: gb_ov:*
9: gb_ov:*
10: gb_ov:*
11: gb_ov:*
12: gb_ov:*
13: gb_ov:*
14: gb_ov:*
15: gb_ov:*
16: gb_ov:*
17: gb_ov:*
18: gb_ov:*
19: gb_ov:*
20: gb_ov:*
21: gb_ov:*
22: gb_ov:*
23: gb_ov:*
24: gb_ov:*
25: gb_ov:*
26: gb_ov:*
27: gb_ov:*
28: gb_ov:*
29: gb_ov:*
30: gb_ov:*
31: gb_ov:*
32: gb_ov:*
33: gb_ov:*
34: gb_ov:*
35: gb_ov:*
36: gb_ov:*
37: gb_ov:*
38: gb_ov:*
39: gb_ov:*
40: gb_ov:*
41: gb_ov:*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3857	100.0	3857	9 HUMERCCAC1	L77890 Homo sapien
2	2866	74.3	2881	9 HS064315	U64315 Human DNA r
3	2718	70.5	2718	6 AR183109	AR183109 Sequence
4	2718	70.5	2718	6 AR203246	AR203246 Sequence
5	2718	70.5	2718	6 AX277006	AX277006 Sequence
6	1932.2	50.1	2858	10 AB017635	AB017635 Citicellu
7	1886	48.9	3445	10 AF189285	AF189285 Mus muscu
8	1872	48.5	3235	10 BC026792	BC026792 Mus muscu
9	1829.4	47.4	35641	6 AX332256	AX332256 Sequence
10	1829.4	47.4	35641	6 AX409578	AX409578 Sequence
11	1829.4	47.4	35641	6 HUMERCCAC1	L77890 Homo sapien
12	1829.4	47.4	119172	9 AC010401	AC010401 Homo sapi
13	1826.2	47.3	149849	9 AC009173	AC009173 Homo sapi
14	1233.4	32.0	29755	9 AF491814	AF491814 Homo sapi
15	1101	28.5	1101	6 AR183108	AR183108 Sequence
16	1101	28.5	1101	6 AR203245	AR203245 Sequence
17	1101	28.5	1101	6 AX277004	AX277004 Sequence
18	803.4	20.8	846	11 G67515	G67515 csnpercc4-P
19	602.4	15.6	718	11 G67516	G67516 csnpercc4-P
20	464.6	12.0	167804	2 AC099214	AC099214 Rattus no
21	459.8	11.9	128026	10 AC004155	AC004155 Mus muscu
22	458.2	11.9	167329	2 AC096307	AC096307 Rattus no
23	448.4	11.6	182509	2 AC117889	AC117889 Rattus no
24	220.4	5.7	312	11 G67513	G67513 csnpercc4-P
25	220.4	5.7	623	9 HSA334998	AJ334998 Homo sapi
26	220.4	5.7	689	9 HSA324979	AJ324979 Homo sapi
27	220.4	5.7	700	9 HSA335478	AJ335478 Homo sapi
28	199.4	5.2	306	11 G67514	G67514 csnpercc4-P
29	171.4	4.4	2064	8 AF191494	AF191494 Arabidops
30	171.4	4.4	3151	8 AF089003	AF089003 Arabidops
31	171.4	4.4	3173	8 AF160500	AF160500 Arabidops
32	159	4.1	160986	3 AC108481	AC108481 Drosophill
33	157.4	4.1	294218	3 AE003430	AE003430 Drosophill
34	157.4	4.1	5123	2 AC014388	AC014388 Drosophill
35	154.6	4.0	3857	2 HUMERCCAC1	L77890 Homo sapien
36	154.6	4.0	29755	9 AF491814	AF491814 Homo sapi
37	154.6	4.0	35641	6 AX332256	AX332256 Sequence
38	154.6	4.0	35641	6 AX409578	AX409578 Sequence
39	154.6	4.0	35641	6 HUMERCCAC1	L77890 Homo sapien
40	154.6	4.0	119172	9 AC010401	AC010401 Homo sapi
41	154.6	4.0	149849	9 AC009173	AC009173 Homo sapi
42	154.6	3.9	182509	2 AC117889	AC117889 Rattus no
43	151.4	3.9	4048	8 AF277377	AF277377 Arabidops
44	150.8	3.9	4048	8 AB010072	AB010072 Arabidops
45	148.8	3.9	83698	8	

ALIGNMENTS

RESULT 1
HUMERCCAC1 3857 bp DNA linear PRI 22-MAR-1997
LOCUS HUMERCCAC1
DEFINITION Homo sapiens excision repair protein ERCC4 mRNA, complete cds,
clone cer4-40.
ACCESSION L77890
VERSION L77890.1 GI:1905923
KEYWORDS
SOURCE Homo sapiens
ORGANISM Homo sapiens adult DNA.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE
1 (bases 1 to 3857)
Brookman,K.W., Lamerdin,J.E., Thelen,M.P., Hwang,M., Reardon,J.T.,
Sancar,A., Zhou,Z.Q., Walter,C.A., Parls,C.N. and Thompson,L.H.

TITLE ERCC4 (XPF) encodes a human nucleotide excision repair protein with
 JOURNAL Mol. Cell. Biol. 16 (11), 6553-6562 (1996)
 MEDLINE 97042484
 PUBMED 8887684
 REFERENCE 2 (bases 1 to 3857)
 AUTHORS Lamerdin, J. E.
 TITLE Direct Submission
 JOURNAL Submitted (03-DEC-1996) J. E. Lamerdin, Human Genome Center,
 Lawrence Livermore National Laboratory, 7000 East Ave, Livermore,
 CA, USA, 94551 jamee@egc.llnl.gov owetornak.llnl.gov
 COMMENT GDSB:S:109760
 FEATURES
 source
 1. 3857
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /map="16p13.13"
 /cell_line="Hela"
 /cell_type="fibroblast"
 /dev_stage="adult"
 13. 2763
 /gene="ERCC4"
 13. 2763
 /note="ERCC4"
 /note="putative"
 /codon_start=1
 /product="DNA repair protein"
 /protein_id="AA050174.1"
 /db_xref="GI:1905924"
 /db_xref="GDB:119113"
 /translation="MESGOPARIAMAPLLEYEROLVELLDLTDLVLCARGLADRL
 LKFLQHLCHPACLVLYLNTOPAEVEYFINOKIEVEHLPRVTEITSNSRYEYV
 OGCVFATSRILVDFDLDRIPSDLTGLVYRAHIIISCOEAFILRLPRKNKGF
 IKAFTDNAAVPTGCHVERVMNLEFRKLYLMPRHVAVNSELEQKREVEYIHSM
 TPLMALIOTAILDIINACLKELCHNPSEVEDLSLEMTIGPEPKTIRHYDPLMHO
 LGAKTSIVODIKILRTLOYLSOYDCYFNLBESLRTAKRQNSWMLDSDTS
 METNARARYNHPDAMSKREKISEKMEIKESBEIKKELVLSNPKWEKLTVEKTE
 AENKESALGPGQVYLICASDRTCSQLDYITLGEAFLLRLYRTFERKDSKAEEW
 MKFRKDESSKRIKSHKRPDKONKERRASTKRLTKKRLKLLTOMVGKPELEBEG
 DVEEGYRREIVSSPEEIKHEEFDVNLSSDAAGLKEPLTIHPLKPELSDPYAL
 TVLHEVERRYVLYDAELTFVROLEIVASRPGRARVYFLIGGSTEORLYLAR
 KEKAEKILIRKASVNVPEERERDETLDLVRGASADVSTDPKAGSOGSONQO
 STVDMREKRSPLSLIHRGDIETLEVGDYILTPMCKYERKSIDLSGSLNKG
 LISOCSMSKRIKRYVLLIEFDPSPKFSITSGALFOELSSDISSKLTLLHLHPRL
 RLWCSPHATLELEFEELKQSPQDPAATAIATDSETLPESEKYNPQDYLKMP
 GVNANCRSLMHVKNIAELALSDDELTSILGNANAKOLYDFIHTSPAEVSKGK
 KR"
 3812
 /note="putative"
 3845
 /note="putative"
 polyA_site
 BASE COUNT 1134 a 851 c 850 g 1022 t
 ORIGIN
 Query Match 100.0% Score 3857: DB 9: Length 3857:
 Best Local Similarity 100.0% Pred. No. 0:
 Matches 3857: Conservative 0: Mismatches 0: Indels 0: Gaps 0:

QY 241 AAGTAGAGAGGTGAACACCTCCCGCGGTGTACAAATGAAATCACAGACAGT 300
 Db 241 AAGATGAGAGAGGTGAACACCTCCCGCGGTGTACAAATGAAATCACAGACAGT 300
 QY 301 CGGTATGAGTATACACAGAGGTGTATATTTGCGCAAGTATGATCTGTGT 360
 Db 301 CGGTATGAGTATACACAGAGGTGTATATTTGCGCAAGTATGATCTGTGT 360
 QY 361 GACTTCTTGACTGATAGAAATCTTACAGATTAAATTAAGTGCATCTTGGTATAGAGCC 420
 Db 361 GACTTCTTGACTGATAGAAATCTTACAGATTAAATTAAGTGCATCTTGGTATAGAGCC 420
 QY 421 CACAGATTAATCGAGTGTTCAGAGAACATCATCTGGCGCTTTCGCCGAAAAAC 480
 Db 421 CACAGATTAATCGAGTGTTCAGAGAACATCATCTGGCGCTTTCGCCGAAAAAC 480
 QY 481 AAGCTGTGTTTATTAAGCTTTCACAGAACATGCTTGTGCTTGTATGATGTTTGT 540
 Db 481 AAGCTGTGTTTATTAAGCTTTCACAGAACATGCTTGTGCTTGTATGATGTTTGT 540
 QY 541 CATGTGAAAGAGTATGAGAAATCTTTTGTGAGGAACTGTATCTGTGCCAAGTTC 600
 Db 541 CATGTGAAAGAGTATGAGAAATCTTTTGTGAGGAACTGTATCTGTGCCAAGTTC 600
 QY 601 CATGTGACACTAAACATCTTTTGAACACACAAACCTGACATGTTAGAAATCATGTT 660
 Db 601 CATGTGACACTAAACATCTTTTGAACACACAAACCTGACATGTTAGAAATCATGTT 660
 QY 661 TCTATGACACCTACCATGCTGCTGTATACAGACTGTATATGTGACATTTTAAATGCATGT 720
 Db 661 TCTATGACACCTACCATGCTGCTGTATACAGACTGTATATGTGACATTTTAAATGCATGT 720
 QY 721 CTAAAGAACTAAAGCCATAACCATGCTGAAGTGAAGATTAATCTTAAGAAAT 780
 Db 721 CTAAAGAACTAAAGCCATAACCATGCTGAAGTGAAGATTAATCTTAAGAAAT 780
 QY 781 GCTATTGAAACCTTTTGAACAAGACAATCGCATATCTGTGATCTTTGTGGACACAG 840
 Db 781 GCTATTGAAACCTTTTGAACAAGACAATCGCATATCTGTGATCTTTGTGGACACAG 840
 QY 841 CTGTGAGCCAAAGCTAATCTTACAGATTTTGAAGATTAATGACACTTTGTGCGAC 900
 Db 841 CTGTGAGCCAAAGCTAATCTTACAGATTTTGAAGATTAATGACACTTTGTGCGAC 900
 QY 901 TATCTCTCAGATGATGTTGTGCACATTTCTTAATCTTGGAAATCTGAGAGCAAG 960
 Db 901 TATCTCTCAGATGATGTTGTGCACATTTCTTAATCTTGGAAATCTGAGAGCAAG 960
 QY 961 GAAAAAGCTTTGCTCAGAAATGAGTTGGCTTTCTTGACATCCAGACACCTGATGTTT 1020
 Db 961 GAAAAAGCTTTGCTCAGAAATGAGTTGGCTTTCTTGACATCCAGACACCTGATGTTT 1020
 QY 1021 ATAAATGCTGACAGAGGTTTATCATCTCCAGATGCCAAATGAGTAAAGAAAAA 1080
 Db 1021 ATAAATGCTGACAGAGGTTTATCATCTCCAGATGCCAAATGAGTAAAGAAAAA 1080
 QY 1081 ATATCTGAAAAAATGAAAAATTAAGAGGAGAAAAAAGCAACTGCTCTAGAA 1140
 Db 1081 ATATCTGAAAAAATGAAAAATTAAGAGGAGAAAAAAGCAACTGCTCTAGAA 1140
 QY 1141 AGCAACCCAAAGTGGAGGAGCACTGAGTAAGTATTAAGAAATTTAGCGCAGAAAAATAG 1200
 Db 1141 AGCAACCCAAAGTGGAGGAGCACTGAGTAAGTATTAAGAAATTTAGCGCAGAAAAATAG 1200
 QY 1201 GAGAGTGAAGCTTGTGCTGCTCAGAGTCAAGTCAATTTGTGCAAGTGTGACCAACA 1260
 Db 1201 GAGAGTGAAGCTTGTGCTGCTCAGAGTCAAGTCAATTTGTGCAAGTGTGACCAACA 1260
 QY 1261 TGTTCACAGCTGAGAGACTATATCACTCTTGGAGCGGAGGCTTCTTATTTAGAGCTCTAC 1320
 Db 1261 TGTTCACAGCTGAGAGACTATATCACTCTTGGAGCGGAGGCTTCTTATTTAGAGCTCTAC 1320

QY	1321	AGAAAAACCTTTGAGAAAGATAGCAAAGCTGAAAGAAAGTGTGATGAAATTTAGAAAGAA	1380
Db	1321	AGAAAAACCTTTGAGAAAGATAGCAAAGCTGAAAGAAAGTGTGATGAAATTTAGAAAGAA	1380
QY	1381	GACACTTCAAGAGAAATTAGAAATCTCAAAAAGACTAAAGACCCCAAAACAAAGAA	1440
Db	1381	GACACTTCAAGAGAAATTAGAAATCTCAAAAAGACTTAAGACCCCAAAACAAAGAA	1440
QY	1441	CGGGCTTCTACCAAGAAAGAACCTTCAAAAAGAAAAAGAGTTGACCTTAACTA	1500
Db	1441	CGGGCTTCTACCAAGAAAGAACCTTCAAAAAGAAAAAGAGTTGACCTTAACTA	1500
QY	1501	ATGGTAGGAAAAACCTGGAAGAACTGGAAGAGAAAGGAGATGTGAGAGAAAGATATGTGCA	1560
Db	1501	ATGGTAGGAAAAACCTGGAAGAACTGGAAGAGAAAGGAGATGTGAGAGAAAGATATGTGCA	1560
QY	1561	GAATAATAGCACTACCCAGAAAGCTCCCGAAGAAATTAACATGAAAGAAATTTGATGTA	1620
Db	1561	GAATAATAGCACTACCCAGAAAGCTCCCGAAGAAATTAACATGAAAGAAATTTGATGTA	1620
QY	1621	AATTTGTATCGGATGCTGCTTTGGAAATCTGAAAGAAACCCCTCATATCATCCATCG	1680
Db	1621	AATTTGTATCGGATGCTGCTTTGGAAATCTGAAAGAAACCCCTCATATCATCCATCG	1680
QY	1681	CTTCTGGGTTGCACAGACCCCTAATGCTGACAAAGGATCTACATGAAGTGGAGCAAGA	1740
Db	1681	CTTCTGGGTTGCACAGACCCCTAATGCTGACAAAGGATCTACATGAAGTGGAGCAAGA	1740
QY	1741	TACGTGTTCTTTATGACAGCAGAGCTAACCTTTGTTGGGACGCTTGAATTTACAGGCG	1800
Db	1741	TACGTGTTCTTTATGACAGCAGAGCTAACCTTTGTTGGGACGCTTGAATTTACAGGCG	1800
QY	1801	AGTAGCCCTGGGAAACCTCGAGAGGTTTCTTATATAGGAGGTCAACGAGGA	1860
Db	1801	AGTAGCCCTGGGAAACCTCGAGAGGTTTCTTATATAGGAGGTCAACGAGGA	1860
QY	1861	CAACGCTATCTCACTGCTTGGCGAAAAAGAAAGAAAGCTTTTGA AAAA ACTCATAGGGA	1920
Db	1861	CAACGCTATCTCACTGCTTGGCGAAAAAGAAAGAAAGCTTTTGA AAAA ACTCATAGGGA	1920
QY	1921	AAAGCAAGCATGTTGTCCTGAAAGAAAGGAGAGATGAACAACAACCTTAGACCTA	1980
Db	1921	AAAGCAAGCATGTTGTCCTGAAAGAAAGGAGAGATGAACAACAACCTTAGACCTA	1980
QY	1981	GTAAGAGGCACAGCATCTGCAGATGTTTCCACTGTGACACTCGGAAAGCCGCTGGCAGGA	2040
Db	1981	GTAAGAGGCACAGCATCTGCAGATGTTTCCACTGTGACACTCGGAAAGCCGCTGGCAGGA	2040
QY	2041	CAGAAATGCTACACAGCAAAACATAGTTTGATATGCGTGAATTTGCAAGTATGACTTCA	2100
Db	2041	CAGAAATGCTACACAGCAAAACATAGTTTGATATGCGTGAATTTGCAAGTATGACTTCA	2100
QY	2101	TCTCTGATTCATCGTCGGGCGCATTTGACATTTGAACCCGCTACTTTAGAGTTGGAGATTAC	2160
Db	2101	TCTCTGATTCATCGTCGGGCGCATTTGACATTTGAACCCGCTACTTTAGAGTTGGAGATTAC	2160
QY	2161	ATCTCTCACTCCAGAAATGTGCGTGAGGCCAAGATATCAGATTAATGCGCTTTTA	2220
Db	2161	ATCTCTCACTCCAGAAATGTGCGTGAGGCCAAGATATCAGATTAATGCGCTTTTA	2220
QY	2221	AATTAAGCGCGCTTACACACAGTGCATCTCATATGTCOCGCTATTAACAACGTCGCGTG	2280
Db	2221	AATTAAGCGCGCTTACACACAGTGCATCTCATATGTCOCGCTATTAACAACGTCGCGTG	2280
QY	2281	CTTCTGATTTGAGTTTGACCTTAGAAGCCTTTCTCTCACTTCCGAGGTGCTTGT	2340
Db	2281	CTTCTGATTTGAGTTTGACCTTAGAAGCCTTTCTCTCACTTCCGAGGTGCTTGT	2340
QY	2341	CAGAGATTCACAGCATGATTAATTTCAAAAGTCACTTTCTTACACTTCCCTCC	2400
Db	2341	CAGAGATTCACAGCATGATTAATTTCAAAAGTCACTTTCTTACACTTCCCTCC	2400
QY	2401	AGACTTAAGATTTCTGTTGCCCTCTCCCTCAATGCAAGCGGCGAGTTGTTGAGAGCTG	2460

Db	2401	AACTACGAGATTCTCTGGTGGCCCTCTCTCATGCAACGGCGAGGTGTGTTGAGAGCTG	2460
Qy	2461	AAACAAACGAAGCCACACACCTGATGGGGGCGACAGCACTGGCCATTTCACCAATCTGTGA	2520
Db	2461	AAACAAACGAAGCCACACACCTGATGGGGGCGACAGCACTGGCCATTTCACCAATCTGTGA	2520
Qy	2521	ACCCCTTCCGAGTCAGAGAAATATAATCTGAGCCCAAGACTCTCTGTGTAAATAATGGCA	2580
Db	2521	ACCCCTTCCGAGTCAGAGAAATATAATCTGAGCCCAAGACTCTCTGTGTAAATAATGGCA	2580
Qy	2581	GGGGTGAAATGCCAAAAACTGCGCTCCTTGATGCAACAGTTAAGAACATGCGAGAAATTA	2640
Db	2581	GGGGTGAAATGCCAAAAACTGCGCTCCTTGATGCAACAGTTAAGAACATGCGAGAAATTA	2640
Qy	2641	GCAGCCCTGTCAACAAGAGACCTCAGAGATTTCTGGGGAAATCTGCAATGCGCAACAG	2700
Db	2641	GCAGCCCTGTCAACAAGAGACCTCAGAGATTTCTGGGGAAATCTGCAATGCGCAACAG	2700
Qy	2701	CTTTATGATTTTCATTCACACCTCTTTTGGAGAAGTGGTATCAAAAGAAAGGAAAGAAAG	2760
Db	2701	CTTTATGATTTTCATTCACACCTCTTTTGGAGAAGTGGTATCAAAAGAAAGGAAAGAAAG	2760
Qy	2761	TGAACAGATGGCTGTTTTCTTATCCCAAGCTGATCTTTTACGGGCTCTTCCCGACA	2820
Db	2761	TGAACAGATGGCTGTTTTCTTATCCCAAGCTGATCTTTTACGGGCTCTTCCCGACA	2820
Qy	2821	CATCATAGCATTAATTAAATATGTTGGTTCATTTTCATTTCTTTTCAATGCTCTTAAAG	2880
Db	2821	CATCATAGCATTAATTAAATATGTTGGTTCATTTTCATTTCTTTTCAATGCTCTTAAAG	2880
Qy	2881	ATTGTACGGTGGACACAGACAGCAAGATCTCTCTGTAACCTGTGCASTTAGCATCACTTG	2940
Db	2881	ATTGTACGGTGGACACAGACAGCAAGATCTCTCTGTAACCTGTGCASTTAGCATCACTTG	2940
Qy	2941	AACTTGCGCTGTGCTGCTTTTCCCTCCCTGACCGCTATAGCGGCGGTGTGACATGTTT	3000
Db	2941	AACTTGCGCTGTGCTGCTTTTCCCTCCCTGACCGCTATAGCGGCGGTGTGACATGTTT	3000
Qy	3001	CTTTTTAAATAGSTTTGTCAAGATCAAGTAAAGTTCTTACAAAGTATATACAAAGATAG	3060
Db	3001	CTTTTTAAATAGSTTTGTGTGATGAGATCAAGTAAAGTTCTTACAAAGTATATACAAAGATAG	3060
Qy	3061	AAACTTTACGTATCTTAACAGATCTCATTTAAGAAAGAAATATGCTTAAGCTGGCATATGA	3120
Db	3061	AAACTTTACGTATCTTAACAGATCTCATTTAAGAAAGAAATATGCTTAAGCTGGCATATGA	3120
Qy	3121	CGGTGCAGGAGGAAAAAGACAGGCAAGAAAGTACCATTTTAACAGTCTCTGTTA	3180
Db	3121	CGGTGCAGGAGGAAAAAGACAGGCAAGAAAGTACCATTTTAACAGTCTCTGTTA	3180
Qy	3181	TCTATGTGCAATATAATTAACAGTCTTAATGTGCACTTAATCCATGCTGCTGGGCTCTGCA	3240
Db	3181	TCTATGTGCAATATAATTAACAGTCTTAATGTGCACTTAATCCATGCTGCTGGGCTCTGCA	3240
Qy	3241	AATCTGGCTTTGCTGTTGTGTGCTGTCGACGCTTGAACGTATGTTGTGTAGAAATCA	3300
Db	3241	AATCTGGCTTTGCTGTTGTGTGCTGTCGACGCTTGAACGTATGTTGTGTAGAAATCA	3300
Qy	3301	TGTTTCTGACCCCTTGTGCTTACAAAGAGACCTTCTGGAACCTGAGAAACATCTCTTTG	3360
Db	3301	TGTTTCTGACCCCTTGTGCTTACAAAGAGACCTTCTGGAACCTGAGAAACATCTCTTTG	3360
Qy	3361	CCATTCCTGACCACTTCTCTATACCAATTTTCTAGCTCCATATCTGTGCTGTGCTG	3420
Db	3361	CCATTCCTGACCACTTCTCTATACCAATTTTCTAGCTCCATATCTGTGCTGTGCTG	3420
Qy	3421	TCTTAAGGAATTTATGAGACCTTCTACTACTAAATTCGAAGACATCTCTCAAAAACCTG	3480
Db	3421	TCTTAAGGAATTTATGAGACCTTCTACTACTAAATTCGAAGACATCTCTCAAAAACCTG	3480
Qy	3481	GTTACTAGTCTTCTATATGACCTTAACATATGATGATTAATTAATTCATTTGTTCA	3540

DB	Accession	Gene	Protein	Length	Source
DB	3481	CTTGACTAGTCTTCTTAATGACCCCTAAACATATGTAGCATATACATTAATTTTCATTGTTCCCA	3540		
QY	3541	AATTAGATATTTTAAAGCAAAATGAATTAACCTGTTTGCAAAAGTTAAATGATGAAGAGCT	3600		
DB	3541	AATTAGATATTTTAAAGCAAAATGAATTAACCTGTTTGCAAAAGTTAAATGATGAAGAGCT	3600		
QY	3601	CTTGAATTTCTCAATTTTGGACATATTACGTCCTCTTAATATGACAGATCCCTAAGTCCA	3660		
DB	3601	CTTGAATTTCTCAATTTTGGACATATTACGTCCTCTTAATATGACAGATCCCTAAGTCCA	3660		
QY	3661	GCTGGCTAGTACAGAGTTTTTTTCAGACCTTCCTGTTTCTCAGCTCTTAATATGACAGATCCCTAAGTCCA	3720		
DB	3661	GCTGGCTAGTACAGAGTTTTTTTCAGACCTTCCTGTTTCTCAGCTCTTAATATGACAGATCCCTAAGTCCA	3720		
QY	3721	ACGAGCATCATATCTCTAGAAATACAACTTAATGGCAGTGCAGCCGAGATCGCAGCACT	3780		
DB	3721	ACGAGCATCATATCTCTAGAAATACAACTTAATGGCAGTGCAGCCGAGATCGCAGCACT	3780		
QY	3781	GCACCCCTGCTGGGCGACAGTGAAGCTTTGTCCTCTATTACAAAAAGAAAAAGAA	3840		
DB	3781	GCACCCCTGCTGGGCGACAGTGAAGCTTTGTCCTCTATTACAAAAAGAAAAAGAA	3840		
QY	3841	AATACAACTTAAGCTCA 3857			
DB	3841	AATACAACTTAAGCTCA 3857			
RESULT 2					
LOCUS	HSU64315	2881 bp	mRNA	Linear	PRI 06-SEP-1996
DEFINITION	Human DNA repair endonuclease subunit 1 (XPF) mRNA, complete cds.				
ACCESSION	U64315				
VERSION	U64315.1				
KEYWORDS	ERCC4; ERCC1; nucleotide excision repair; xeroderma pigmentosum.				
SOURCE	Homo sapiens.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
AUTHORS	1 (bases 1 to 2881)				
	Sijbers,A.M., de Laat,M.L., Ariza,R.R., Bijlgerstaff,M., Wei,Y.,				
	Moggs,J.G., Carter,K.C., Shell,B.K., Evans,E., de Jong,M.C.,				
	Rademakers,S., de Rooij,J., Jaspers,N.G., Hoeljmakers,J.H. and				
	Wood,R.D.				
	Xeroderma pigmentosum group F caused by a defect in a				
	structure-specific DNA repair endonuclease				
	Cell (1996) in press				
	2 (bases 1 to 2881)				
TITLE	Direct Submission				
JOURNAL	Submitted (18-JUL-1996) Biochemistry of Inherited Syndromes Lab.,				
REFERENCE	Imperial Cancer Research Fund, Clare Hall Laboratories, Blanche				
AUTHORS	lane, South Mims, Herts., EN6 3JD, United Kingdom				
JOURNAL	Location/Qualifiers				
TITLE	1. 2881				
FEATURES	/organism="Homo sapiens"				
source	/db_xref="taxon:9606"				
	/chromosome="16"				
	/map="16p13.1-13.2"				
	1. 2881				
	/gene="XPF"				
	16. 2733				
	/gene="XPF"				
	/codon_start=1				
	/product="DNA repair endonuclease subunit"				
	/protein_id="AAB07689.1"				
	/db_xref="GI:1524411"				
	/translation="WAPLEYEROLVELDLDGLVVCARGLADRLYLHFLQJLHCHIE				
	ACLVAVLTQPAEEYETINQKIEGVEHLPRVNETISNSRYEYTGVIATSRISRSR				
	LDVLDTRPSDLITGLIVRAHRIISCEAEFLRLPRKNGKFIKAFIDAAVAFA				
	DYGHCHVERVMRNLEVRKLYLMPRFHVAVNSFLDQHKREYVEIHVSMPTMLAOTATA				
	LDILNACKEIKCHNPSLEVEDSLNENIGKPRDKTIHLYLDPLMHOGATKTSLNOD				
	LKILATLLQYLSQDVCYFLNLDLESIRATEKAFGNSGWLFLDSTSWFIVARARVTH				
	LPLDKMSKERIKSEMEIKGEDEERKKELVLESNPKMEWLTVELEIEMENESRGLG				

[illegible]

Db 661 ACTGCTATACGACATTTTAAATGCAATGCTAAAGACACTAAATGCAATACCCATCG 720
Qy 751 CTGGAAGTGAAGATTATCTTTAGAAAATGCTATTGGAAAACTTTTGACAGACAATC 810
Db 721 CTGGAAGTGAAGATTATCTTTAGAAAATGCTATTGGAAAACTTTTGACAGACAATC 780
Qy 811 CGCATATATCTGATCTTTTGTGGCAGCAGCTTGAGGCAAGACTTAAATCTTAGTTGAG 870
Db 781 CGCATATATCTGATCTTTTGTGGCAGCAGCTTGAGGCAAGACTTAAATCTTAGTTGAG 840
Qy 871 GATTGGAAGATATTAGCAACTTTGTCAGATCTCTCTCAGTATGATTTGTGACATTT 930
Db 841 GATTGGAAGATATTAGCAACTTTGTCAGATCTCTCTCAGTATGATTTGTGACATTT 900
Qy 931 CTGTAATCTCTGGAATCTCTGAGACAACGAAAAAGCTTTTGTGACAAATTCAGTTGG 990
Db 901 CTGTAATCTCTGGAATCTCTGAGACAACGAAAAAGCTTTTGTGACAAATTCAGTTGG 960
Qy 991 CTGTTCTGACTGACGACCTCGATGTTTAAATGCTGAGCAAGGCTTATCATCTT 1050
Db 961 CTGTTCTGACTGACGACCTCGATGTTTAAATGCTGAGCAAGGCTTATCATCTT 1020
Qy 1051 CCAGATGCCAAAATGAGTAAAAAAGAAAAAATATCTGAAAAATGGAATTTAAAGAAAGG 1110
Db 1021 CCAGATGCCAAAATGAGTAAAAAAGAAAAAATATCTGAAAAATGGAATTTAAAGAAAGG 1080
Qy 1111 GAAGAAACAAAAAAGAACTGCTCTAGAAAACCAACCAAGTGGAGGCACTGACTAA 1170
Db 1081 GAAGAAACAAAAAAGAACTGCTCTAGAAAACCAACCAAGTGGAGGCACTGACTAA 1140
Qy 1171 GTATTAAAGAAATTTAGGCGAAAAATTAAGAGAGTGAAGCTCTTGCGTCCAGTCAA 1230
Db 1141 GTATTAAAGAAATTTAGGCGAAAAATTAAGAGAGTGAAGCTCTTGCGTCCAGTCAA 1200
Qy 1231 GTACTGATTTTGCAGATGATGACCGAACAATGTTCCAGCTGAGAGACTATATCACTCTT 1290
Db 1201 GTACTGATTTTGCAGATGATGACCGAACAATGTTCCAGCTGAGAGACTATATCACTCTT 1260
Qy 1291 GGAGGCGAGGCTCTTCTTTTGAAGGCTCTACAGAAAACTTTGAGAAAGATAGCAAACT 1350
Db 1261 GGAGGCGAGGCTCTTCTTTTGAAGGCTCTACAGAAAACTTTGAGAAAGATAGCAAACT 1320
Qy 1351 GAAGAAGTCTGGATGAATTTAGAGAGAGACAGTCTAAAGAGAAATAGCAAACTCTAC 1410
Db 1321 GAAGAAGTCTGGATGAATTTAGAGAGAGACAGTCTAAAGAGAAATAGCAAACTCTAC 1380
Qy 1411 AAAAGACCTAAAGACCCCAAAACAAAGAACGGGCTTTACCAAGAAAAAACCCTCAA 1470
Db 1381 AAAAGACCTAAAGACCCCAAAACAAAGAACGGGCTTTACCAAGAAAAAACCCTCAA 1440
Qy 1471 AAGAAAAAAGGGAAGTATGACCTTAACCTCAATGATGAGAAAACTGAAAGAACTGGAAG 1530
Db 1441 AAGAAAAAAGGGAAGTATGACCTTAACCTCAATGATGAGAAAACTGAAAGAACTGGAAG 1500
Qy 1531 GAAGGAGATGTCGAGAGAGATATGTCGAGAAATTAAGCAGTACCCGAAAGCTGCGCG 1590
Db 1501 GAAGGAGATGTCGAGAGAGATATGTCGAGAAATTAAGCAGTACCCGAAAGCTGCGCG 1560
Qy 1591 GAAGAAATTTAAGCATGAAGATTTTGATGTAATTTGTATCGATGCTGCTTTGCGAATC 1650
Db 1561 GAAGAAATTTAAGCATGAAGATTTTGATGTAATTTGTATCGATGCTGCTTTGCGAATC 1620
Qy 1651 CTGAAAGAAACCCCTCACTATCATCCGCTTGCGGTGAGGAGGACCCCTATCTCTG 1710
Db 1621 CTGAAAGAAACCCCTCACTATCATCCGCTTGCGGTGAGGAGGACCCCTATCTCTG 1680
Qy 1711 ACAAGGATCTACATGATGAGTGAAGCAAGATGCTGTTCTTTAGAGCAGAGATTAAC 1770
Db 1681 ACAAGGATCTACATGATGAGTGAAGCAAGATGCTGTTCTTTAGAGCAGAGATTAAC 1740
Qy 1771 TTTGTTCCGACGCTTGAATTTTACAGGCGAGTAGGCTTGGAACCTCTGAGGCTTTAC 1830
|||||

Db 1741 TTTGTTCCGACGCTTGAATTTTACAGGCGAGTAGGCTTGGAACCTCTGAGGCTTTAC 1800
Qy 1831 TTTCTTATATAGGAGGTTCACTGAGAGACACGCTATCTCACTACCTTTGCGGAAAGAA 1890
Db 1801 TTTCTTATATAGGAGGTTCACTGAGAGACACGCTATCTCACTACCTTTGCGGAAAGAA 1860
Qy 1891 AAGGAGCTTTTGAAGAAACCTATAGGAAAAAAGCAAGCAGTGTGTGCTTGAAGAAAGA 1950
Db 1861 AAGGAGCTTTTGAAGAAACCTATAGGAAAAAAGCAAGCAGTGTGTGCTTGAAGAAAGA 1920
Qy 1951 GAAGGAGAGATGAAGAAACCTTAGACCTAGTAAGAGGACACAGCAATGCAAGATGTTCC 2010
Db 1921 GAAGGAGAGATGAAGAAACCTTAGACCTAGTAAGAGGACACAGCAATGCAAGATGTTCC 1980
Qy 2011 ACTGACACTCGGAAAGCCGGTGGCCAGAGACAGAAATGATACACAGCAAGCATAGTTGTG 2070
Db 1981 ACTGACACTCGGAAAGCCGGTGGCCAGAGACAGAAATGATACACAGCAAGCATAGTTGTG 2040
Qy 2071 GATATGCGTGAATTTGGAAGTGAAGCTTCCATCTCTGATCCATGCTGCGAGCATTGACAT 2130
Db 2041 GATATGCGTGAATTTGGAAGTGAAGCTTCCATCTCTGATCCATGCTGCGAGCATTGACAT 2100
Qy 2131 GAACCCGTGACTTTAGAGGTTGAGATTTACATCTCACTCCAGAAATGCTGCGAGAGCC 2190
Db 2101 GAACCCGTGACTTTAGAGGTTGAGATTTACATCTCACTCCAGAAATGCTGCGAGAGCC 2160
Qy 2191 AAGATATCATGATTTAATGAGGCTTTAATAAGGCGCGCTCTACAGCAGTGCATC 2250
Db 2161 AAGATATCATGATTTAATGAGGCTTTAATAAGGCGCGCTCTACAGCAGTGCATC 2220
Qy 2251 TCATATGCCGCTACTACAAAGCGTCCGCTTCTGATTTGAGTTGAACTTGAACCTAGCAAGCT 2310
Db 2221 TCATATGCCGCTACTACAAAGCGTCCGCTTCTGATTTGAGTTGAACTTGAACCTAGCAAGCT 2280
Qy 2311 TTCTCTCACTTCCCGAGGTCCTGTTTTCAGAGATCTCAGAGATGACATTAATGTTCC 2370
Db 2281 TTCTCTCACTTCCCGAGGTCCTGTTTTCAGAGATCTCAGAGATGACATTAATGTTCC 2340
Qy 2371 AAATCATCTCTTACACTTCACTTCCCAAGACATGAGATTTCTGTTGGCCCTCTCT 2430
Db 2341 AAATCATCTCTTACACTTCACTTCCCAAGACATGAGATTTCTGTTGGCCCTCTCT 2400
Qy 2431 CATGCAAGGCGGAGTCTTTGAGAGCTGAAACAAAGCAAGCCACGCTGATGCGGCG 2490
Db 2401 CATGCAAGGCGGAGTCTTTGAGAGCTGAAACAAAGCAAGCCACGCTGATGCGGCG 2460
Qy 2491 ACAGCACTGCGCATTTACAGACATTTCTGAACCCCTTCCGAGTCAAGAAATATATCT 2550
Db 2461 ACAGCACTGCGCATTTACAGACATTTCTGAACCCCTTCCGAGTCAAGAAATATATCT 2520
Qy 2551 GGTCCCAAGACTTCTGTTAAATGCGAGGCGTGAATGCCAAACGCGCGCTCTTG 2610
Db 2521 GGTCCCAAGACTTCTGTTAAATGCGAGGCGTGAATGCCAAACGCGCGCTCTTG 2580
Qy 2611 ATGCACACGTTTAAAGACATGACAGATTTAGCAGCCCTGTCAACAAGAGAGCTCAAGAT 2670
Db 2581 ATGCACACGTTTAAAGACATGACAGATTTAGCAGCCCTGTCAACAAGAGAGCTCAAGAT 2640
Qy 2671 ATTCTGGGGAATGTCGAAATGCCAAGACGTTTATGATTTCACTCAACCTTTTGCA 2730
Db 2641 ATTCTGGGGAATGTCGAAATGCCAAGACGTTTATGATTTCACTCAACCTTTTGCA 2700
Qy 2731 GAAGTGTATCAAAAGGAAAAAGGAAAAAGTGAAGAGAGGCTGTTCTTATCCCAT 2790
Db 2701 GAAGTGTATCAAAAGGAAAAAGGAAAAAGTGAAGAGAGGCTGTTCTTATCCCAT 2760
Qy 2791 GCTCTACTTTTCAAGGCTCTCTTCCAGACATCAATAGTCAATTAATTAATGTTGG 2850
Db 2761 GCTCTACTTTTCAAGGCTCTCTTCCAGACATCAATAGTCAATTAATTAATGTTGG 2820
Qy 2851 CTATTTCATCTTTTCCAAATGCTTTAATGATTTGACGCTGACAGACAGACCGAG 2904
Db 2821 CTATTTCATCTTTTCCAAATGCTTTAATGATTTGACGCTGACAGACAGATTCAG 2874
|||||

RESULT 3
ARI83109
LOCUS ARI83109 2718 bp DNA linear PAT 20-APR-2002
DEFINITION Sequence 14 from patent US 6340566.
ACCESSION ARI83109
VERSION ARI83109.1 GI:20226702
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 2718)
McClutchen-Maloney,S.L.
AUTHORS Detection and quantitation of single nucleotide polymorphisms, DNA
TITLE sequence variations, DNA mutations, DNA damage and DNA mismatches
JOURNAL Patent: US 6340566-A 14 22-JAN-2002;
FEATURES
source 1..2718
Location/Qualifiers
BASE COUNT 827 a 588 c 634 g 669 t
ORIGIN
Query Match 70.5%; Score 2718; DB 6; Length 2718;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2718; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 46 ATGGCCCCGCTGCTGAGTACGAGCAGCAGCTGCTGTAAGTCTCGACACTGACGGG 105
DB 1 ATGGCCCCGCTGCTGAGTACGAGCAGCAGCTGCTGTAAGTCTCGACACTGACGGG 60
QY 406 CTAGTAGTGTGCGCCCGCGGGCTCGCGCGGACCGGCTCCTTACCACTTCTCCAGCTG 165
DB 61 CTAGTAGTGTGCGCCCGCGGGCTCGCGCGGACCGGCTCCTTACCACTTCTCCAGCTG 120
QY 166 CACTGCCACCCAGCCTGCTGCTGCTGCTGCTCAACACGACGCGCGCGAGAGAGAT 225
DB 121 CACTGCCACCCAGCCTGCTGCTGCTGCTGCTCAACACGACGCGCGCGAGAGAGAT 180
QY 226 TTATCAATCAGCTGGAAGATAGAAAGATTGAACACCTCCGCGCGGTGAACAAATGAA 285
DB 181 TTATCAATCAGCTGGAAGATAGAAAGATTGAACACCTCCGCGCGGTGAACAAATGAA 240
QY 286 ATCACAAGCAAGCTGCTATGAAGTTTACACACAAGGTGTGTATTTATTTGGACAAGT 345
DB 241 ATCACAAGCAAGCTGCTATGAAGTTTACACACAAGGTGTGTATTTATTTGGACAAGT 300
QY 346 AGGATACCTTGTGCTGATCTTCTGACTGATAGATACCTTTCAGATTATTTACTGCGATC 405
DB 301 AGGATACCTTGTGCTGATCTTCTGACTGATAGATACCTTTCAGATTATTTACTGCGATC 360
QY 406 TTGGTGTATAGAGCCCAAGAAATATCGAGTCTTGTCAAGAACATTCATCTTGGCCATC 465
DB 361 TTGGTGTATAGAGCCCAAGAAATATCGAGTCTTGTCAAGAACATTCATCTTGGCCATC 420
QY 466 TTTCCGCCAGAAAACAACAGCTGTTTATTAAGCTTTTACAGACACATGCTGTGCTTT 525
DB 421 TTTCCGCCAGAAAACAACAGCTGTTTATTAAGCTTTTACAGACACATGCTGTGCTTT 480
QY 526 GATACGTGTTTTTGTGATGTGGAAGAGATGATGAGAAATCTTTTGTGAGGAACCTGAT 585
DB 481 GATACGTGTTTTTGTGATGTGGAAGAGATGAGAAATCTTTTGTGAGGAACCTGAT 540
QY 586 CTGTGGCCAAAGTTCATGATAGCACTAATCTTTTATTTAGACAGACAAACCTGAAGTT 645
DB 541 CTGTGGCCAAAGTTCATGATAGCACTAATCTTTTATTTAGACAGACAAACCTGAAGTT 600
QY 646 GTAGAAATCCATGTTTCTATGACACCTACCATGCTTGTCTATAAGACTCTATCTAGAC 705
DB 601 GTAGAAATCCATGTTTCTATGACACCTACCATGCTTGTCTATAAGACTCTATCTAGAC 660
QY 706 ATTTTAATGCACTGTCTAAGAGAACTAAATGCCATACCCATCGTTGAAGTGAAGAT 765
DB ATTTTAATGCACTGTCTAAGAGAACTAAATGCCATACCCATCGTTGAAGTGAAGAT 720

DB 661 ATTTTAATGCACTGTCTAAGAGAACTAAATGCCATACCCATCGTTGAAGTGAAGAT 720
QY 766 TTATCTTTAGAAAATGCTATTTGAAAACCTTTTGACAGACAATCCGCCATTATCTGGAT 825
DB 721 TTATCTTTAGAAAATGCTATTTGAAAACCTTTTGACAGACAATCCGCCATTATCTGGAT 780
QY 826 CTTTGTGGCACCAGCTTGGAGCCAAAGACTAAATCCCTTGTGAGATTTGAAGATATTA 885
DB 781 CTTTGTGGCACCAGCTTGGAGCCAAAGACTAAATCCCTTGTGAGATTTGAAGATATTA 840
QY 886 CGAATCTTGTGCAATATCTCTCAGTATGATTTGTGTCACATTTCTTAATCTTCTGAA 945
DB 841 CGAATCTTGTGCAATATCTCTCAGTATGATTTGTGTCACATTTCTTAATCTTCTGAA 900
QY 946 TCTCTGAGACCAACGAAAAAGCTTTTGTGAGAAATTCAGGTTGCTGTTCTTGACTCC 1005
DB 901 TCTCTGAGACCAACGAAAAAGCTTTTGTGAGAAATTCAGGTTGCTGTTCTTGACTCC 960
QY 1006 AGCAGCTCGATGTTTAAATGCTGAGCAAGGGTTTATCATCTTCACATGCCAAATG 1065
DB 961 AGCAGCTCGATGTTTAAATGCTGAGCAAGGGTTTATCATCTTCAGATGCGCAAAATG 1020
QY 1066 AGTAAAAAGAAAAAATATCTGAAAAAATGAAATTAAGAAAGGAGAGAAACAAAAAG 1125
DB 1021 AGTAAAAAGAAAAAATATCTGAAAAAATGAAATTAAGAAAGGAGAGAAACAAAAAG 1080
QY 1126 GAACTGCTCTAGAAACCAACCAAGTGGAGGCACTGACTGATATTTAAAGAAAT 1185
DB 1081 GAACTGCTCTAGAAACCAACCAAGTGGAGGCACTGACTGATATTTAAAGAAAT 1140
QY 1186 GAGCGAAGAAATPAGAGAGATGAAAGCTTGGTGGTCCAGAGTCAACTACTGATTTTGCA 1245
DB 1141 GAGCGAAGAAATPAGAGAGATGAAAGCTTGGTGGTCCAGAGTCAACTACTGATTTTGCA 1200
QY 1246 AGTAGATGACGAACATGTTCCAGCTGAGAGACTATATCATCTTGTGAGCGAGGCTTC 1305
DB 1201 AGTAGATGACGAACATGTTCCAGCTGAGAGACTATATCATCTTGTGAGCGAGGCTTC 1260
QY 1306 TTATTTAGGCTCTACAGAAAACTTTGAGAAAGATAGCAAAAGCTGAGAGTGTGATG 1365
DB 1261 TTATTTAGGCTCTACAGAAAACTTTGAGAAAGATAGCAAAAGCTGAGAGTGTGATG 1320
QY 1366 AAATTTAGAGAGAGAGACAGTTCAAAGAAATTAGAAATCTCACAAAGACTTAAGAG 1425
DB 1321 AAATTTAGAGAGAGAGACAGTTCAAAGAAATTAGAAATCTCACAAAGACTTAAGAG 1380
QY 1426 CCCCCAAAACAAGAGAGGCTTCTACCAAAAGAAAGAACCTCAAAAGAAAAAGGAG 1485
DB 1381 CCCCCAAAACAAGAGAGGCTTCTACCAAAAGAAAGAACCTCAAAAGAAAAAGGAG 1440
QY 1486 TTGACCTTAACCAATTTGTAGAGAAACCTGAAAGAACTGGAAGAGAGAGATGTGAG 1545
DB 1441 TTGACCTTAACCAATTTGTAGAGAAACCTGAAAGAACTGGAAGAGAGAGATGTGAG 1500
QY 1546 GAAGATATCTGAGAGAAATTAAGCACTAGCCAGAAAGCTGCCGAGAAATTAAGAT 1605
DB 1501 GAAGATATCTGAGAGAAATTAAGCACTAGCCAGAAAGCTGCCGAGAAATTAAGAT 1560
QY 1606 GAAAGATTTGATGATTAATTTGTATCGATGCTGCTTGTGGAATCTGAAAGAACCCCTC 1665
DB 1561 GAAAGATTTGATGATTAATTTGTATCGATGCTGCTTGTGGAATCTGAAAGAACCCCTC 1620
QY 1666 ACTATCATCATCCGCTTGTGAGTGGAGAGACCCCTATGCTGCAAGAGGATAGAT 1725
DB 1621 ACTATCATCATCCGCTTGTGAGTGGAGAGACCCCTATGCTGCAAGAGGATAGAT 1680
QY 1726 GAAAGTGAGCAAGATACCTGCTTTTATGACGAGAGACTAACCCTTTGTTGGAGCTT 1785
DB 1681 GAAAGTGAGCAAGATACCTGCTTTTATGACGAGAGACTAACCCTTTGTTGGAGCTT 1740
QY 1786 GAAATTTACAGGGCAGATAGGCTGGAAGAACTCTGAGGCTTACCTTTCTTATACGGA 1845
DB 1741 GAAATTTACAGGGCAGATAGGCTGGAAGAACTCTGAGGCTTACCTTTCTTATACGGA 1800

QY	886	CGAATTTGCTGAGTATCTCTCAGTATGATTTGTGACATTTCTTAATCTTCTGAA	945
Db	841	CGAATTTGCTGAGTATCTCTCAGTATGATTTGTGACATTTCTTAATCTTCTGAA	900
QY	946	TCTCTGAGAGCAAGAAAAAGTTTGTGAGAAATTCAGTTGGCTTTCTTGACACC	1005
Db	901	TCTCTGAGAGCAAGAAAAAGTTTGTGAGAAATTCAGTTGGCTTTCTTGACACC	960
QY	1006	AGCAGCTCGATGTTTATTAATGCTCGAGCAAGGTTTATCATCTTCCAGATGCCAAATG	1065
Db	961	AGCAGCTCGATGTTTATTAATGCTCGAGCAAGGTTTATCATCTTCCAGATGCCAAATG	1020
QY	1066	AGTAAAAAGAAAAATATCTGAAAAATTTAAGAAAGGGAAGAAAAAAG	1125
Db	1021	AGTAAAAAGAAAAATATCTGAAAAATTTAAGAAAGGGAAGAAAAAAG	1080
QY	1126	GAATGCTCTAGAAAGCAACCAAGTGGAGGCACTGACTGATGAATTTAAAGAAAT	1185
Db	1081	GAATGCTCTAGAAAGCAACCAAGTGGAGGCACTGACTGATGAATTTAAAGAAAT	1140
QY	1186	GAGCAGAAAAATAGAGAGTGAAGCTTTGGTCCAGTCAAGTACGATTTGTGCA	1245
Db	1141	GAGCAGAAAAATAGAGAGTGAAGCTTTGGTCCAGTCAAGTACGATTTGTGCA	1200
QY	1246	AGTATGACGCAACATGTTCCAGCTGAGAGACTATATCACTCTTGGAGGCGAGCCCTTC	1305
Db	1201	AGTATGACGCAACATGTTCCAGCTGAGAGACTATATCACTCTTGGAGGCGAGCCCTTC	1260
QY	1306	TTATTTAGGCTCTACAGAAAAACCTTTGAGAGGATAGCAAAAGCTGAAGATGTGATG	1365
Db	1261	TTATTTAGGCTCTACAGAAAAACCTTTGAGAGGATAGCAAAAGCTGAAGATGTGATG	1320
QY	1366	AAATTTAGGAAGAAAGAGTTCAAAGAAATTTGAAATCTCCAAAGACCTAAAGAC	1425
Db	1321	AAATTTAGGAAGAAAGAGTTCAAAGAAATTTGAAATCTCCAAAGACCTAAAGAC	1380
QY	1426	CCCCAAAACAAGAGAGGGCTTCTACCAAGAAAGAAACCCCTCAAAAAGAAAAAGGAG	1485
Db	1381	CCCCAAAACAAGAGAGGGCTTCTACCAAGAAAGAAACCCCTCAAAAAGAAAAAGGAG	1440
QY	1486	TTGACCTTAATCTCAATGATGTAAGAAACCTGAAAGACTGGAAGGAAAGATGTGAG	1545
Db	1441	TTGACCTTAATCTCAATGATGTAAGAAACCTGAAAGACTGGAAGGAAAGATGTGAG	1500
QY	1546	GAAAGATATGCTGAGAAATTAAGCAGTAGCCAGAAAGCTGCCGGAAGAAATTAAGCAT	1605
Db	1501	GAAAGATATGCTGAGAAATTAAGCAGTAGCCAGAAAGCTGCCGGAAGAAATTAAGCAT	1560
QY	1606	GAAAGATTTGATGTAATTTGTTCATCGGATGCTGTTGGAAATCTGAAAGAACCCCTC	1665
Db	1561	GAAAGATTTGATGTAATTTGTTCATCGGATGCTGTTGGAAATCTGAAAGAACCCCTC	1620
QY	1666	ACTATCATCATCCGCTTTCGGGTTGAGAGGAGCCCTATGCTGCAAGAGGATCTATCAT	1725
Db	1621	ACTATCATCATCCGCTTTCGGGTTGAGAGGAGCCCTATGCTGCAAGAGGATCTATCAT	1680
QY	1726	GAAATGAGGCAAGATACGTGTTCTTTATGACGAGCTAAACCTTTGTTGGCAGCTT	1785
Db	1681	GAAATGAGGCAAGATACGTGTTCTTTATGACGAGCTAAACCTTTGTTGGCAGCTT	1740
QY	1786	GAAATTTACAGAGGCGATAGGCTGGGAAACCTCTGAGGGTTTACTTTTATATACGA	1845
Db	1741	GAAATTTACAGAGGCGATAGGCTGGGAAACCTCTGAGGGTTTACTTTTATATACGA	1800
QY	1846	GCTTCAACTGAGGAAACAGCTATCTCTGCTTTGGGAAAGAAAGAAAGGCTTTGAA	1905
Db	1801	GCTTCAACTGAGGAAACAGCTATCTCTGCTTTGGGAAAGAAAGAAAGGCTTTGAA	1860
QY	1906	AAACTCATAGGAAAAAGCAAGCATGTTGCCCTGAAGAAAGAAAGGAGAGATGAA	1965
Db	1861	AAACTCATAGGAAAAAGCAAGCATGTTGCCCTGAAGAAAGAAAGGAGAGATGAA	1920
QY	1966	ACAACTTAGACCTTAGTAAGAGGACAGCATCTGCAATGTTTCCATCTGACACTCGAA	2025
Db	1921	ACAACTTAGACCTTAGTAAGAGGACAGCATCTGCAATGTTTCCATCTGACACTCGAA	1980
QY	2026	GCCGCTGGCCAGAAACAGAAATGATACAGCAAGACATAGTTGTGCAATACGGAATTT	2085
Db	1981	GCCGCTGGCCAGAAACAGAAATGATACAGCAAGACATAGTTGTGCAATACGGAATTT	2040
QY	2086	CGAATGAGCTTCCATCTCTGATATCATGCTGGGCAATTCATTAACACCGTACATTTA	2145
Db	2041	CGAATGAGCTTCCATCTCTGATATCATGCTGGGCAATTCATTAACACCGTACATTTA	2100
QY	2146	GAGTTGAGATTTACATCCCTACCTCCAGAAATGGCGGGAGCCGCAAGATAGATGAT	2205
Db	2101	GAGTTGAGATTTACATCCCTACCTCCAGAAATGGCGGGAGCCGCAAGATAGATGAT	2160
QY	2206	TTAATCGCTCTTTAAATACGGCCGCTCTACAGCCAGTGCATCTCCATGCTCCGCTAC	2265
Db	2161	TTAATCGCTCTTTAAATACGGCCGCTCTACAGCCAGTGCATCTCCATGCTCCGCTAC	2220
QY	2266	TACAAGGCTCCGCTGCTTCTGATTTGAGTTGACCTTAGCAAGCCTTTCTCTCACTTCC	2325
Db	2221	TACAAGGCTCCGCTGCTTCTGATTTGAGTTGACCTTAGCAAGCCTTTCTCTCACTTCC	2280
QY	2326	CGAGTGCCTTGTTCAGAGATCTCCAGCAATGACATTTGTTCCAAACTCACTCTTCT	2385
Db	2281	CGAGTGCCTTGTTCAGAGATCTCCAGCAATGACATTTGTTCCAAACTCACTCTTCT	2340
QY	2386	ACACTTCACTTCCCGACACTAGAGATTTCTGTGTCCTCTCTCTCATGCAAGCGCGAG	2445
Db	2341	ACACTTCACTTCCCGACACTAGAGATTTCTGTGTCCTCTCTCTCATGCAAGCGCGAG	2400
QY	2446	TTGTTTGAAGAGCTGGAACAAAGCAAGCAAGCAGCTGATGGGGAGACAGATGCGCAT	2505
Db	2401	TTGTTTGAAGAGCTGGAACAAAGCAAGCAAGCAGCTGATGGGGAGACAGATGCGCAT	2460
QY	2506	ACAGCAGATTTCTGAACCCCTTCCGAGTCAAGAGATATCTCTGCTCCCAAGACTTC	2565
Db	2461	ACAGCAGATTTCTGAACCCCTTCCGAGTCAAGAGATATCTCTGCTCCCAAGACTTC	2520
QY	2566	TTGTTTAAAAATGCCAGGGGTGAATGCCAAAAACCTGCCCTTTGATGACACAGCTTAAG	2625
Db	2521	TTGTTTAAAAATGCCAGGGGTGAATGCCAAAAACCTGCCCTTTGATGACACAGCTTAAG	2580
QY	2626	AACATGCGAGATTTAGAGCCCTGCTCAAGAGAGCTCACAGATTTTCGGGAAATGCT	2685
Db	2581	AACATGCGAGATTTAGAGCCCTGCTCAAGAGAGCTCACAGATTTTCGGGAAATGCT	2640
QY	2686	GCAAAATGCCAAACAGCTTTATGATTTTCATTCACACCTCTTTTGCAGAAATCGATCAAAA	2745
Db	2641	GCAAAATGCCAAACAGCTTTATGATTTTCATTCACACCTCTTTTGCAGAAATCGATCAAAA	2700
QY	2746	GGAAGAGGAAAAAGTGA 2763	
Db	2701	GGAAGAGGAAAAAGTGA 2718	
RESULT 5			
AX277006			
LOCUS			
DEFINITION			
AX277006			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
human.			
Homo sapiens			
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.			
REFERENCE			
1			
McCutchen-Maloney, S.L.			
Chimeric proteins for detection and quantitation of dna mutations,			
dna sequence variations, dna damage and dna mismatches			
Patent: WO 01/3079-A 14 04-OCT-2001;			
JOURNAL			

The Regents of The University of California (US)

FEATURES

Location/Qualifiers

1..2718

/organism="Homo sapiens"

/db_xref="taxon:9606"

BASE COUNT 827 a 588 c 634 g 669 t

ORIGIN

Query Match 70.5%; Score 2718; DB 6; Length 2718;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 2718; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

46 ATGGCGCCGCTCTGGAGTAGAGGACAGCTGTGCTGGAAGTGGTCGACACTGACGAGG 105
1 ATGGCGCCGCTCTGGAGTAGAGGACAGCTGTGCTGGAAGTGGTCGACACTGACGAGG 60
106 CTAGTAGTGTGGCG 165
61 CTAGTAGTGTGGCG 120
166 CACTGCACCCAGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 225
121 CACTGCACCCAGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 180
226 TTTATCAATCAGCTGAAGTAGAGGAGTTGAACACCTCCCTGCGCGGTAACTAAATGAA 285
181 TTTATCAATCAGCTGAAGTAGAGGAGTTGAACACCTCCCTGCGCGGTAACTAAATGAA 240
286 ATCACAGCAGCAGTGTGCTATGAGTTTACACAGAGGTGTATATTTTGGCAGCAAT 345
241 ATCACAGCAGCAGTGTGCTATGAGTTTACACAGAGGTGTATATTTTGGCAGCAAT 300
-346 AGATATCTGTGGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 405
301 AGATATCTGTGGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 360
406 TTGCTGTATAGAGCCACAGAAATTAATCGAGTCTGTCAAGAGCAATCATCTTGCCTTC 465
361 TTGCTGTATAGAGCCACAGAAATTAATCGAGTCTGTCAAGAGCAATCATCTTGCCTTC 420
466 TTTCCGCGAGAAAACAAAGCTGTTTATTTAAAGCTTTACAGACATCTGTTGCTTTT 525
421 TTTCCGCGAGAAAACAAAGCTGTTTATTTAAAGCTTTACAGACATCTGTTGCTTTT 480
526 GATATCGTGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 585
481 GATATCGTGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540
586 CTGTGGCCAGAGTTCATGATAGCAGTAAATCTATTTTGAACAGCAGCAAACTGAAAT 645
541 CTGTGGCCAGAGTTCATGATAGCAGTAAATCTATTTTGAACAGCAGCAAACTGAAAT 600
646 CTAGAAATTCATGTTTCTATGACACCTACATGCTTGTCTATACAGCTCTTACTGGAC 705
601 GTAGAAATTCATGTTTCTATGACACCTACATGCTTGTCTATACAGCTCTTACTGGAC 660
706 ATTTAAATGCATGCTCTAAAGAGCTAAATATCCATTAACCCATCCGTTAAAGTGAAGAT 765
661 ATTTAAATGCATGCTCTAAAGAGCTAAATATCCATTAACCCATCCGTTAAAGTGAAGAT 720
766 TTTATCTTAAAGAAATGCTATTTGAAAAACCTTTTGAACAGACATCCGCTTATATCGAT 825
721 TTTATCTTAAAGAAATGCTATTTGAAAAACCTTTTGAACAGACATCCGCTTATATCGAT 780
826 CTTTGTGGCAGCAGCTTGGAGCCAAAGCTAAATCTTATGTTCAAGATTTTGAAGATTA 885
781 CTTTGTGGCAGCAGCTTGGAGCCAAAGCTAAATCTTATGTTCAAGATTTTGAAGATTA 840
886 CGAATTTGCTGAGATATCTCTCAGATGATGATGTCACATTTCTTATATCTTCTGAAA 945
841 CGAATTTGCTGAGATATCTCTCAGATGATGATGTCACATTTCTTATATCTTCTGAAA 900
946 TCTGTAGAGCAACGAAAAAGCTTTTGTGAGAAATTCAGTTGGCTGTTCTTGACTCC 1005

|||||
Db 901 TCTGTAGAGCAACGAAAAAGCTTTTGTGAGAAATTCAGTTGGCTGTTCTTGACTCC 960
1006 AGCACTGCATGTTTATTAATGCTCGACAGAGGCTTATCATCTTCCAGATGCCAAAATG 1065
961 AGCACTGCATGTTTATTAATGCTCGACAGAGGCTTATCATCTTCCAGATGCCAAAATG 1020
1066 AGTAAAAAGAAAAATATCTGAAAAATGAAATTAAGAAAGGGGAGCAACAAAAAG 1125
1021 AGTAAAAAGAAAAATATCTGAAAAATGAAATTAAGAAAGGGGAGCAACAAAAAG 1080
1126 GAACCTGCTCTAGAAAAGCAACCCAAAGTGGAGGACCTGACTGAACTATTTAAAGAAAT 1185
1081 GAACCTGCTCTAGAAAAGCAACCCAAAGTGGAGGACCTGACTGAACTATTTAAAGAAAT 1140
1186 GAGGCAAAATTAAGAGAGTGAAGCTCTTGTGTGTCAGAGTCAATCTGATTTTGCA 1245
1141 GAGGCAAAATTAAGAGAGTGAAGCTCTTGTGTGTCAGAGTCAATCTGATTTTGCA 1200
1246 AGTATGACCGAAACATGTTCCAGCTGAGAGACTATATCATCTTGTGAGGAGGCTTC 1305
1201 AGTATGACCGAAACATGTTCCAGCTGAGAGACTATATCATCTTGTGAGGAGGCTTC 1260
1306 TATTGAGGCTCTACAGAAAACCTTTGAGAAAGATAGCAAAAGCTGAGAGTCTGATG 1365
1261 TATTGAGGCTCTACAGAAAACCTTTGAGAAAGATAGCAAAAGCTGAGAGTCTGATG 1320
1366 AAATTTAGAGAGAGACAGTTCAAAAGAGATTTAGAAATCTTCACAAAAGACCTAAAGAC 1425
1321 AAATTTAGAGAGAGACAGTTCAAAAGAGATTTAGAAATCTTCACAAAAGACCTAAAGAC 1380
1426 CCCCAGAAAGAAAGAGGCTCTTACAAAAGAAAGAACCTCAAAAAGAAAAAGGAAAG 1485
1381 CCCCAGAAAGAAAGAGGCTCTTACAAAAGAAAGAACCTCAAAAAGAAAAAGGAAAG 1440
1486 TTGACCTTAACCTCAATATGTTAGAGAAAACCTGAAAGACTGGAAGAGATGTGAGAG 1545
1441 TTGACCTTAACCTCAATATGTTAGAGAAAACCTGAAAGACTGGAAGAGATGTGAGAG 1500
1546 GAAGATATGCTGAGAAATTAAGCACTAGCCAGAAAGCTGCCGCAAGAAATTAAGCAT 1605
1501 GAAGATATGCTGAGAAATTAAGCACTAGCCAGAAAGCTGCCGCAAGAAATTAAGCAT 1560
1606 GAAGATTTGATGTAATTTGATGATGAGATGCTGCTTGGAAATCTGAAAGAACCCCTC 1665
1561 GAAGATTTGATGTAATTTGATGATGAGATGCTGCTTGGAAATCTGAAAGAACCCCTC 1620
1666 ACTATCATCATCCGCTTGTGGTTGACAGGACCCCTATGCTCTGCAAGAGGTACTACAT 1725
1621 ACTATCATCATCCGCTTGTGGTTGACAGGACCCCTATGCTCTGCAAGAGGTACTACAT 1680
1726 GAAGTGAGCCAAAGATAGCTGTTCTTTATGACGCAAGAGCTAATCTTGTGGCAGCTT 1785
1681 GAAGTGAGCCAAAGATAGCTGTTCTTTATGACGCAAGAGCTAATCTTGTGGCAGCTT 1740
1786 GAAATTTACAGGCGAGTGTGGCCCTGGGAAACCTCTAGGGTTACTTCTTATATCGGA 1845
1741 GAAATTTACAGGCGAGTGTGGCCCTGGGAAACCTCTAGGGTTACTTCTTATATCGGA 1800
1846 GGTTCAGTGAAGAACAGCTATCTCACTGCTTTCGGAAGAAAGAAAGAGCTTTTGA 1905
1801 GGTTCAGTGAAGAACAGCTATCTCACTGCTTTCGGAAGAAAGAAAGAGCTTTTGA 1860
1906 AAATCTATAAGGAAAAACCAAGCATGTTGTCCGAAAGAAAGAAAGCAAGATGTA 1965
1861 AAATCTATAAGGAAAAACCAAGCATGTTGTCCGAAAGAAAGAAAGCAAGATGTA 1920
1966 ACAAACTTAGACCTAGTAGAGGACAGCATGCTGAGATGTTTCCATGACACTCGGAA 2025
1921 ACAAACTTAGACCTAGTAGAGGACAGCATGCTGAGATGTTTCCATGACACTCGGAA 1980
2026 GCCGCTGCGCAGAACAGAAATGTTACAGCAAGCAATGATGTTGCAATATGCGTGAATTT 2085
|||||

Db 1981 GCCGGTGCCAGGAAGAATGTAACAGCAAGATAGTTGTGATATGCGTAATTT 2040
 QY 2086 CGAAGTAGCTTCATCTGTATGATCGTCGGGGCATTTGACATTTGAACCCGTACTTTA 2145
 Db 2041 CGAAGTAGCTTCATCTGTATGATCGTCGGGGCATTTGACATTTGAACCCGTACTTTA 2100
 QY 2146 GAGGTTGAGATTTACATCTCTACTCCAGAAATGTGCGTGAGCCGAGAGATATCATGAT 2205
 Db 2101 GAGGTTGAGATTTACATCTCTACTCCAGAAATGTGCGTGAGCCGAGAGATATCATGAT 2160
 QY 2206 TTAATCGGCTTTTAATTAACGGCCGCTTACAGCCAGTGCATCTCCCGCTAC 2285
 Db 2161 TTAATCGGCTTTTAATTAACGGCCGCTTACAGCCAGTGCATCTCCCGCTAC 2220
 QY 2266 TACAAAGGTCGCCGCTTGTATGATTTGACCTTACAGCAAGCTTCTCTCACTTCC 2325
 Db 2221 TACAAAGGTCGCCGCTTGTATGATTTGACCTTACAGCAAGCTTCTCTCACTTCC 2280
 QY 2326 CGAAGTCCCTTGTTCAGAGATCTCCAGCAATGACATTAATTTCAAACTCACTTCTT 2385
 Db 2281 CGAAGTCCCTTGTTCAGAGATCTCCAGCAATGACATTAATTTCAAACTCACTTCTT 2340
 QY 2386 ACACCTACTTCCCAAGCACTACGATTTCTGTGCTCCCTCTCTCATGCAAGCGGAG 2445
 Db 2341 ACACCTACTTCCCAAGCACTACGATTTCTGTGCTCCCTCTCTCATGCAAGCGGAG 2400
 QY 2446 TTGTTTGAAGAGCTGGAACAAAGCAAGCAAGCCTGATGGCGGACAGCACTGGCATT 2505
 Db 2401 TTGTTTGAAGAGCTGGAACAAAGCAAGCAAGCCTGATGGCGGACAGCACTGGCATT 2460
 QY 2506 ACACGACATTTCTGAACCCCTTCCGAGTCAAGAGATTAATCTGTGCTCCCAAGACTTC 2565
 Db 2461 ACACGACATTTCTGAACCCCTTCCGAGTCAAGAGATTAATCTGTGCTCCCAAGACTTC 2520
 QY 2566 TTGTTTAAATATGCCAGGGGTGAATGCCAAATGCCGCTCTTGATGACACAGCTTAAG 2625
 Db 2521 TTGTTTAAATATGCCAGGGGTGAATGCCAAATGCCGCTCTTGATGACACAGCTTAAG 2580
 QY 2626 AACATCCAGATTTAGCAGGCTGTCAAGAGCAGCTCAGATTTCTGGGAAATGCT 2685
 Db 2581 AACATCCAGATTTAGCAGGCTGTCAAGAGCAGCTCAGATTTCTGGGAAATGCT 2640
 QY 2686 GCAATATCCCAAGCCTTTATGATTTCAATTCACACCTCTTTTGCAAGAGTGTATCAAA 2745
 Db 2641 GCAATATCCCAAGCCTTTATGATTTCAATTCACACCTCTTTTGCAAGAGTGTATCAAA 2700
 QY 2746 GGAAGAGGGAAGAGTGA 2763
 Db 2701 GGAAGAGGGAAGAGTGA 2718
 RESULT 6
 LOCUS AB017635 2858 bp mRNA linear ROD 08-JAN-2000
 DEFINITION Cricetus griseus mRNA for ERCC4, complete cds.
 ACCESSION AB017635
 VERSION AB017635.1 GI:6683475
 KEYWORDS ERCC4.
 SOURCE Cricetus griseus CHO cell_line:CHO-9 cDNA to mRNA.
 ORGANISM Cricetus griseus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
 Cricetus.
 REFERENCE 1 (bases 1 to 2858)
 AUTHORS Hayashi,T.
 TITLE CHO ERCC4 cDNA
 JOURNAL Published Only in Database (2000)
 REFERENCE 2 (bases 1 to 2858)
 AUTHORS Hayashi,T.
 JOURNAL Direct Submission
 Submitted (11-SEP-1998) Tsuyuko Hayashi, Institute of Development,
 Aging and Cancer Tohoku University, Molecular Genetics, 4-1
 Seiry-machi, Aoba-ku, Sendai, Miyagi 980-77, Japan

(E-mail: tsuyuidag.tohoku.ac.jp, Tel:81-22-717-8469,
 Fax:81-22-717-8470)
 FEATURES
 source
 Location/Qualifiers
 1..2858
 /organism="Cricetus griseus"
 /db_xref="taxon:10029"
 /cell_line="CHO-9"
 /cell_type="CHO"
 1..2858
 /gene="ERCC4"
 43..2784
 /gene="ERCC4"
 /codon_start=1
 /product="ERCC4"
 /protein_id="BA89229.1"
 /db_xref="GI:6683475"
 translation="MDGISAVRKAMAPLLEVEROLVELLDGLVVCARGLGADRL
 LYLHLRHCHPACLVLVNTPAEEEFINOLKEVEHLEPRVTEITSNREYVT
 OGGLIFATSRILVDFLDRIPDLITGLIVRAHRIEISQEAFLIRLPRKNKRGF
 IKAFTDVAVDFDGFCHVERVMNLFVRKLYMIRKRVAVNSFLDQHPVLEHVM
 TPAMISTOTLIDILINACLEIKCHNSLEVEDLSFNALGKRPDKIRIYLDPIHIO
 LKAKTSLVODLILRLILOYSQIDCVTFELNLESIRATEKYPGONSLFLDAST
 MEVNARVRVPDVKLNKAKSESSEGEETKELVESNPMEALSEVLKETIEN
 KESBALGPGQVILICASDRTCCQLRDYLTAGAVFLRLRYRTEPKDSAAEYVNL
 RKSDGRKTRKSDKRPKDTNKRERASTKPKRKRELTLQVMGAEKPEEGAAE
 EDQORATSSPEGGEIHEAFDNLNLSDSAYGLKEPLTIHPLVGCSDPYALRV
 LHEVEPRVVLVYDAELTFVRLQLEIRASRPGKRLRYFLIYGGSTEOERYLTLRREK
 EAEKELIREKASVYPERGREDTLCYRIYSDAPADTKAGKQDNRGNQPSIV
 YDKRERSELPILHRGIDIEVYTLLEVGYIILPELCVERKYSDLTGLNSGRITS
 QCLAMSRVYRRPVLLIEFDAGKRPESLAPRSFQESSSDVSSKLTILHLPRRLIL
 WCPSPHATAELEFLRQNKPOPDAAVTAMATTAOSETLIPESDKNPGQDPLVLMPCIN
 AKNCHSIENHVKNIKIAELASLSLGHAGNANVOLYDFLHTAVADYVSGGRVVR"
 BASE COUNT 759 a 722 c 776 g 601 t
 ORIGIN
 Query Match 50.1%; Score 1932.2; DB 10; Length 2858;
 Best Local Similarity 81.5%; Pred. No. 0;
 Matches 2253; Conservative 0; Mismatches 503; Indels 9; Gaps 1;
 Db 1 GGAAGAGCTTCATGAGTCAAGGAGACCGGCTCGAGGAGTTCGCTGGCGGCTGCTG 60
 QY 1 GGAAGAGCTTCATGAGTCAAGGAGACCGGCTCGAGGAGTTCGCTGGCGGCTGCTG 60
 Db 31 GGAAGAGCTTCATGAGTCAAGGAGACCGGCTCGAGGAGTTCGCTGGCGGCTGCTG 98
 QY 61 GAGTACGAGCAGACAGCTGTGCTGGAACCTGCTGCACTGACGGCTAGTATGTGGCC 120
 Db 91 GAGTACGAGCAGACAGCTGTGCTGGAACCTGCTGCACTGACGGCTAGTATGTGGCC 150
 QY 121 CGCGGGCTCGGCGGAGACCGGCTCTTACACCTTTCTCCAGCTGCACCTGCCACGCC 180
 Db 151 CGCGGGCTCGGCGGAGACCGGCTCTTACACCTTTCTCCAGCTGCACCTGCCACGCC 210
 QY 181 TGCTGTGCTGTGCTGCACACCGCAGCCGCGCAGAGAGATTTTATTAATCACTGTG 240
 Db 211 TGCTGTGCTGTGCTGCACACCGCAGCCGCTGAGGAGATTTTATTAATCACTGTG 270
 QY 241 AAGATAGAGAGTGAACACCTCCCTGCGCGTGTAAACAATTAATGACAAAGCAAGT 300
 Db 271 AAGATAGAGAGTGAACACCTCCCTGCGCGTGTAAACAATTAATGACAAAGCAAGT 330
 QY 301 CGCTATGAAGTTTACACACAAGAGTGTATTAATTTGCGACAGTACATCTGTGGTT 360
 Db 331 CGCTATGAAGTTTACACACAAGAGTGTATTAATTTGCGACAGTACATCTGTGGTT 390
 QY 361 GACTTCTTGACTGTAAATATACCTTCAGATTTATTTACTGGCATCTGTGTATAGACC 420
 Db 391 GATTTCCTTGACTGTAAATATACCTTCAGATTTATTTACTGGCATCTGTGTATAGACC 450
 QY 421 CACAGATTAATCGAGCTGTGTCAGAGCAAGCAATTAATCTTGGCCCTTTGGCAGAAAC 480
 Db 451 CACAGATTAATCGAGCTGTGTCAGAGCAAGCAATTAATCTTGGCCCTTTGGCAGAAAC 510
 QY 481 AAGCGTGTATTAAGCTTTCACAGACATCTGTGCTTTCATCTGATGATTTTGT 540

D	b	511	AAGCGTGGCTTTCATCAAAAGCTTTACAGACAACGCTGtgcctttttgacacTggtttttgt	570
Q	y	541	CATGTGGAAGAGTGTATGAGAAATCTTTTTgTgAGAAACGTGTATCTGTGCCAAGTTC	600
D	b	571	CATGTGGAAGAGTGTATGAGAAACCTTTTTgTgAGAAAGCTTACTGTGGCCAAAGTTC	630
Q	y	601	CATGTACCAATAACATCTTTTAgAAAGCAACCAACCTGAAGTGTGAATAATCATGT	660
D	b	631	CATGTACCAATCAACTGTCTTTTgTgAAACAACAACCAAGCTGGAAGTGTGAATAATCAGTg	690
Q	y	661	TCTATGACACTTACACTGTCTTGTATACAGACTGCTATCTGTGACATTTTAAATGCAGT	720
D	b	691	TCTATGACTCTCGCATGCTTTCATTCACAGACGCAATCTAGCATCTTAAATGCGGT	750
Q	y	721	CTAAAGAACTAAATATGCCATTAACCATTCGCTTGAAGTGAAGATTTATCTTTGAATAAT	780
D	b	751	CTGAAGAACTGAAGTGGCCCAACCCATCATAGAAAGTGAAGACTTGTCTACTGAAAT	810
Q	y	781	GCTATGGAAGAACTTTTgTgCAAGAACATCCGCATATATCTGATCTTGTGGACACG	840
D	b	811	GCTTTTGGGAAGCCGTTTgTgCAAGAACATCCGCATTACTTGTGACCTTTTATGGCACG	870
Q	y	841	CTTGGAGCCAAAGACTAAATCTTAgTTCAGAGATTTGAAGATTTTGCAGACTTGTGCAG	900
D	b	871	CTTGGAGCCAAAGACTAAATCTTGTGACAGAGATCTGAAGATCTTAAGAACCTTACTGAG	930
Q	y	901	TATCTCTCATGATGATGTTGTGTACATATTTCTTAACTCTTGTGAATCTCTGAGACAAG	960
D	b	931	TATCTCTCATGATGATGTTGTGTATATTTCTTAACTCTTGTGAATCTCTGAGACAACA	990
Q	y	961	GAAGAACTTTTGTGTGAGATTCAGTGTGGCTTCTTGAATCAGACCTGCATGTTT	1020
D	b	991	GAGAAAGTCTTTTGTGATTCAGATTCAGTGTGGCTTCTTGTGATCAGACCTGCATGTTT	1050
Q	y	1021	ATTAATCTCTGAGCAGGTTTTCATCTTCCAGATGCCAAATGATTAATAAAGAAAA	1080
D	b	1051	GTTGATCTCTGTGCACAGAGTTTTCGTGTCCCGATGTCAAACTGACAAAAAGGCCAA	1110
Q	y	1081	ATATCTAAAAAAATGCAATTAAGAAAGGGGAAAGCAACAAAAAGAACTGGTCTAGAA	1140
D	b	1111	ATGTCT-----GAGAGTGCAGAAAGGGCAAGAAACAAAAAGAACTGGTCTAGAA	1161
Q	y	1141	AGCAACCCCAAGTGGGAGAGGACACTGATGAAGTTTAAAAAGAAATTGAGSCAGAAATPAG	1200
D	b	1162	AGCAACCCCAAGTGGGAGGAGGATTTAGTGAAGTCTGAAAGAAATGGAAGCAAGAAATPAG	1221
Q	y	1201	GAGAGTGAACCTTGTGTGTCCAGGTCAAGTACTGATTTGCAAGTGAATGACCGAACA	1260
D	b	1222	GAGAGTGAACCTTGTGTGTGTCCAGGCGAGGTGTGATGTGTGCAAGGATGACCGACT	1281
Q	y	1261	TGTTCCAGTGAAGACTATCTACTGTTGAGCGGAGGCTTCTTATGAGGCTTAC	1320
D	b	1282	TGCTGCAGGTGAGAGTACTCTGACCGCAGAGAGCAAGAGCTTCTCTCGGCTTAC	1341
Q	y	1321	AGGAAACCTTGAAGAGATAGCAAGCTGAAAGAGTGTGAATTTAGAAAGAA	1380
D	b	1342	AGGAAACCTTGAAGAGAGCAACCAAGCGCAAGAGTGTGGTGAATTTAGAAAGGCG	1401
Q	y	1381	GACAGTTCAAGAGAAATTTAGAAATTCACAAAGACTTAAAGACCCCCCAACAAAGAA	1440
D	b	1402	GAGCGTCCAAAGAGAAACATGAATCTGACAAAGACCCAAAGACCAACAAAGAAAG	1461
Q	y	1441	CGGGCTTCTACCAAGAGAACCTTCAAAAAGAAAAAGSAGTGTGACTTAACTCA	1500
D	b	1462	CGTGCCTCTACCAAAAAGGGGGCTCTTAAAGGAAAGAACCGAGACTTACTGTGACGCA	1521
Q	y	1501	ATGTAGAAAGAACTGTGAAGTGAAGAGAGAGATGTGAGAAAGAAATATGTGGA	1560
D	b	1522	GTGATGGGAGACTGCTGAGAGCCACCAAGAGAGAGGGGCTGTGACAGAGAAAGACAGAGGA	1581
Q	y	1561	GAATTAAGCAGTAGCCCAAGAAAGTCCCGGAGAGAAATTAAGCATGAAGAAATTTGATGA	1620
D	b	1582	CAACCCACAGTAGTGTCCAGAGAGGCTCGGGGGAGAGATTCACATGAAGCTTTGATTTA	1641

QY	1631	AAATTGATCGGATGCTGCTTTCGGAAATCTGAAAGAACCCCTCAATTATCATCCATCG	1680
Db	1642	AAATTGTCATCTGATCGATCGGGCTTAATGCGATCTGAAAGAACCCCTCACCATATCATCCG	1701
QY	1681	CTTCGGGTTGCACCGACCCCTAATGCGTGCAGCAAGGGTCTCAATCAAGTGGAGCCAGA	1740
Db	1702	CTTGTGGGGTGCACCGACCCCTTAATGCTTGTACGGGGGTACTCAATCAAGTGGAGCCAGA	1761
QY	1741	TACGTGTTCTTTATATGACGACGAGACTAACTTGTGTGGCAGCTTGAATAATTACAGGGCG	1800
Db	1762	TACGCGTGTCTGTATGATGACGAGCTACACTTGTGTACGGCAGCTGCATCTACAGGGCC	1821
QY	1801	AGTAGGCGCTGGGAAACCTCTGAGGGTTTACTTTCTTATATACGAGNGTTCAACTGAGGA	1860
Db	1822	AGCAGGCCACGGGAAACCACTGAGGGTTTACTTTCTTATATGAGGGTCTCAACGGAAAGAA	1881
QY	1861	CAACGCTATCTCACTGCTGTTTGGGAAAGAAAGAAAGCTTTTGAAAGAACTCATAGGGAA	1920
Db	1882	CAGCGTTAATCGACCGCTCTTTCGAAAGAAAGAAAGAGCATTTTGAAAGACTCATAGGGAA	1941
QY	1921	AAAGCAAGCATGTTTGTTCCTTGAAAGAAAGAGAGAGATGAACAACTTAGACCTA	1980
Db	1942	AAAGCTAGCATGTTTGTTCCTTGAAAGAGAGAGAGATGAACCAACTGAGACTTA	2001
QY	1981	GTAGAGGGCCAGCATCTGAGATGTTTCCACTGACATCGGAAGCCGAGCGGCCAGGAA	2040
Db	2002	GCAAGAGGCCACAGTGTCCACAGATGCTCCCGCTGACATCGGAAGAAAGTGGTGGCCAGAG	2061
QY	2041	CAGATGGTATACACGCAAGCAACATAGTTGTGGATATCGTGAATTTCGAAAGAGCTTCCA	2100
Db	2062	CATTAACGGCACACAACCCAGCATTTGTGGTGGATCATGGAGATTTTCGGAGTAGCTACCA	2121
QY	2101	TCTCTGATCCATGCTCGGGGCAATTGACATTTGAACCCGCTACTTTAGAGTTGGAGATTAC	2160
Db	2122	TCCCTGATCCATCGAGGGGCAATGACATCGACCCAGTACTCTGAGAGTGGGTGATTAAC	2181
QY	2161	ATTCCTACTCCAGAAAGATGCTCGTGGAGCGCAGAGATGATAGATTATATCGGCTCTTTA	2220
Db	2182	ATTCGAGACGCTGAGTGTGTGTGGAGCGCAAAAGCGTAGTAGACTCATCGGCTCGCTGA	2241
QY	2221	AATTAACGGCGCCCTTACACGAGTGCATCTCATGTCCCGCTACTACAAAGCGTCCGCTG	2280
Db	2242	AATAGTGGCCGCTGTACACCAAGTGTCTAGCCATATCAAGCTACTACCGGGGCAAGTA	2301
QY	2281	CTTCTGATGTAGTTGTGACCCCTAGCAAGCCCTTCTCTCACTTCCGAGGTGCCCTGTGTT	2340
Db	2302	CTGCTCATACGAGTTCCGATGCTGGTAAGCCCTTCTCTTGGCGCCACGGGTTCTTCTTC	2361
QY	2341	CAGGAGATCTTCACGCAATATGACATTAGTTCGAAACGTCGTTCTTACACTTCACTTCCCG	2400
Db	2362	CAGGAGATGATATACACGACGAGCTGAGCTCCAAACCTCAGCCTCTCACTCAGCTTCCCG	2421
QY	2401	AGACAGAGGATTTCTGTGGTCCCGCTCCCTCAATGACAAAGCGGAGATCTGTTGAGGAGCTG	2460
Db	2422	GCGCTGGGGCTGTCTGTGGTCCCGCTCCCTCAATGACAAAGCGGAGCTGTTGAGGAGCTG	2481
QY	2461	AAACCAAGCAAGCCACAGCCTGATATGGCGGCACAGCACTGGCCATTACAGCAATCTTGAA	2520
Db	2482	AAGCAAGCAAGCCACAAACCTGATATGACGACGAGCGGCACTCACTGCAAGACTCGGAA	2541
QY	2521	ACCTTCCCGAGTACAGAAAGTATATCTGTGTTCCCAAGACTTCTTGTAAAAATGCGCA	2580
Db	2542	ACGCTGCAGAGTGTGGCAAGTATTAACCCCGGCCCCCAAGACTTGTGTAAAAAGTGTCT	2601
QY	2581	GGGGTGAATGGCAAAACGCGCGCTCTTATGTCACACAGTTTAAACAACTCCACAATTTA	2640
Db	2602	GGGATCATTGGCAAGAACTGTCACTCCCTGTATTAATACAGTAGACATCTAGCAGAGCTG	2661
QY	2641	GCAGCCCTGTCAAGACAGAGCTCACGAGTATTTCTGGGGAATGCTGCAAAATGCCAAACAG	2700
Db	2662	GCCTCCCTGTGTCCAGAGAGCGGCTACAGCATCTCTGGGCAATGCGGGAATATCCCAAGAC	2721

QY	2701	CTTTATGATTTTCATTCACACCTCTTTTGGCAGAAAGTCGATCCAAAGGAAAGGAAAG	2760
Db	2722	CTGTACGATCTTCTCCACACGTGCTATGACAGACGTGTGTCGGAGGACGAGTGTGA	2781
QY	2761	TGAC	2765
Db	2782	TGACC	2786
RESULT	7		
LOCUS	AF189285	3445 bp	mRNA
DEFINITION	Mus musculus DNA repair endonuclease Xpf mRNA, complete cds.		
VERSION	AF189285		
KEYWORDS	AF189285.1	GI:6063148	
SOURCE	Mus musculus.		
ORGANISM	Mus musculus		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
AUTHORS	1 (bases 1 to 3445) Shannon,M., Lamerdin,J.E., Richardson,L., McCutchen-Maloney,S.L., Hwang,M.H., Handel,M.A., Stubbs,L. and Thelen,M.P.		
TITLE	Characterization of the mouse Xpf DNA repair gene and differential expression during spermatogenesis		
JOURNAL	Genomics 62 (3), 427-435 (1999)		
MEDLINE	20112756		
REFERENCE	10644440		
AUTHORS	2 (bases 1 to 3445) Shannon,M. and Thelen,M.P.		
TITLE	Direct Submission		
JOURNAL	Submitted (23-SEP-1999) Molecular and Structural Biology Division, Lawrence Livermore National Laboratory, L452, P.O. Box 808, Livermore, CA 94550, USA		
FEATURES	Location/Qualifiers		
source	1..3445		
	/organism="Mus musculus"		
	/db_xref="taxon:10090"		
	/chromosome="16"		
	/map="proximal to Apod"		
	/cell_type="pachytene spermatocyte"		
	200..2953		
	/function="remove DNA damage in nucleotide excision repair pathway"		
	/note="5' incision endonuclease"		
	/codon_start=1		
	/product="DNA repair endonuclease Xpf"		
	/protein_id="AA03157.1"		
	/db_xref="GI:6063149"		
	/translation="MERGLSGERSMADILEYEROVLEILDSNGLVAVCARGLCTDRILYHRLRHCPACIVLYLNQPAEEFIMQKLEGYEHLPRTNTNIAASREYVTTGGTITVSRILVDFLGTIRPSDLITLVYRAHRLIESQDEAFILRLFRKKRKEFIKATFDAAVAFDGFCHVERVMNRLFVRKLYLWRFIVAVNSFLEQKPEVEIVHSMTCMAIOTIATLILDLINACLEKHCNPSELEVEDSLNALGKPEDKTIRHLDPLMHQLGAKTKSLVODLKITPLTLQYSDCVTFLLNLESLRATKEVVGOMSGMLFDASISAMFNARAYRVDPDVLTKAKTSEKTSPEVOBTKEIVLESNPKMEALIDVLEKLEIEAKNESFALVGGPGVILTKCASSDRTCCOLRLYLSAGARTFILLRLYRKFEDGKAEEYVAVNRKGGPKRTTSDKRRPAAPKKEASAKRQCPILRKKOELTLYVUSAEPPEDKALBEDICROTSSSPESCGVEIKRESDFLWVSSAAAGILKEPLTIHPLIGCDPYA LTRVLEVERRYVVLXDDELTFVQLEITYVASRPGKLRVYFLILYGGSTEORYLALARKKEAEFEKILIREKASVYVEERGBRETNIDLARGSAALDAPDTRKAGQEQENGTOSSIVLDEKRESELPSLIHRRGIDIEPVTLVEGDYILTPELCVERKESVDLIGSLHSGRRLVSOCILASGRYRRPVLILEPDSPKFSILAPRGAFQFOMSSVSSKLTITLHPFRRLRLMGCSPHATAELEBELKONRQPPAATAMATTADESILPESDRNRPQPDVILKMRPGVNAKKCRSLMNOVKNIATIELATLSTELRTLLTIGHSNAKOLHDFLHTAYADLVSKGRVRK."		
BASE COUNT	852 a	913 c	916 g
ORIGIN			764 t
Query Match	48.9%	Score 1886	DB 10; Length 3445;
Best Local Similarity	80.4%	Prod. No. 0;	
Matches 2232; Conservativity	0;	Mismatches 540;	Indels 3; Gaps 1;

OY	1	GGAGAGCTTCATGGAGTCAAGGACGCCGGCTGCAGCGATTGCCATGCCGCCGCTGCTG	60
Db	188	GGAAAGGGGGCCCTATGGAGCCAGGCGCTGTGCGGCGAAGCGGGGAGTATGGCGCGCTGTGG	247
OY	61	GAGTACGAGCGCACACTGGTGTGTGAACCTGTCTGCAGACTGCAGCGGCTAGTACTGTGCC	120
Db	248	GAGTACGAGCGCACAGAGTGTGTGAACCTGTCTGCAGACTGCAGCGGCTAGTACTGTGTGCC	307
OY	121	CGCGGGCTCGGGCGGAGCGGCTCCTCAACACTTTCACAGCTGCAGTGCACCCAGCC	180
Db	308	CGCGGACTCGGAGCCGAGCGGCTCCTCAACACTTTCAGAGTGCAGCGGCTAGTACTGTGTGCC	367
OY	181	TGCTGTGTGTGTGTGTCTCAACACGACGCCGCCGAGAGAGTATTTATCAATACAGT	240
Db	368	TGCTGTGTGTGTGTGTCTCAACACGACGCCGCCGAGAGAGTATTTATCAATACAGT	427
OY	241	AAGATGAGAGGATGGAACACCTCCCTCGCGGTGTAAACAATGAAATCAACAGCAACGT	300
Db	428	AAGATGAGAGGATGGAACATCTCCCTCGAGAGTGCAGCAATGAAATGCGCACTAACGT	487
OY	301	CGCTTTGGAAGTTTACACACAAAGGTGTGTATATTTGGCAGCAAGTAAAGTACTGTGGTT	360
Db	488	CGCTATGAAGTGTACACGAGGAGGTGTATATTTTGAACAACAGCGGATACTGCTGGTT	547
OY	361	GACTTCTTGACTGATAGANTACCTTCAGATTTAACTTACTGGCATCTGTGTATAGACC	420
Db	548	GATTTCTTGACAGGTAGATATCTTCAGATTTAACTGGGCATCCGTGTACAGGGCA	607
OY	421	CACACAATAATGSACTGTGTGAABAAGCATCTTGGCGCTCTTGCGCAGAAAAAC	480
Db	608	CACACAATCATTTGACTCTGCGCAGAGCTTCATCTGCGCTCTCTCCGCCAGAAAAAC	667
OY	481	AAACCTGGTTTTATTTAAAGCTTTTCACAGACATGCTGTGCTCTTGAATAGTGGTTTTGT	540
Db	668	AAGCGGGGTTTCATAAAGCTTTTCACCGACAAAGCTGTGTGCTTTCACACTGGTTTTGT	727
OY	541	CATGTGGAAGAGTATGAGAAATCTTTTGTGAGAAACTGTATCTGTGGCAAGTTTC	600
Db	728	CACGTGGAAGAGTATGATGCGAAACCTTTTGTGAGAAAGCTCTTACCTGTGGCAAGTTTC	787
OY	601	CATGAGGAGTAACTCATTTTAAACAGCAACCAACCGAAAGTGTAAAGAAATCATGTT	660
Db	788	CATGAGGAGTAACTCATCTTGTGAGACAAACAAGCCGTGAAGTGTATAGAGATTCAGTGG	847
OY	661	TCTATGACACCTACACCTCTTGTCTTACAGATGCTATACAGACATTTTAAATCATGT	720
Db	848	TCTATGACTCCCGCATGCTTGTCCATTCAGAGCGGCATACTGACATCTTAAATGCAATGT	907
OY	721	CTAAAGAACTTAAATGCCATTAACCCATGCTTGAAGTGAAGATTTATCTTAGAAAT	780
Db	908	CTGAAGAACTTAAATGTCAACACCCTGCTACTTGAAGTGAAGATTTGTCACTTAAGAAAC	967
OY	781	GCTATTGAAAACTTTGACAAAGCAATCGCGCATTTATCTGGATCTTTGGGAGCAG	840
Db	968	GCTATTGGAAGCCATTTGACAAAGCAATCGCGCATTTACTTGGACCTTTGGGAGATCG	1027
OY	841	CTTGGAGCCAGACTTAAATCCTTACTTACAGATTTTGAAGATTTTACGAATTTGCTGCAG	900
Db	1028	CTTGGAGCCAGACTTAAATCCTTGTGTCAGATTTTAAAGATTTCTAAGAACTTACTGCAG	1087
OY	901	TATCTCTCTCAGTATGATTTGTGTACATTTCTTAACTTTCTGGAATTTCTGTAGACCAAG	960
Db	1088	TATCTCTCTCAGTATGATTTGTGTACATTTCTTGAATTTCTGTGGAATTTCTGTAGACCA	1147
OY	961	GAAAAAGCTTTTGTGTCAGAAATTCAGGTGGCTTTTCTGTGATGTCTGACACCTTCATGTTT	1020
Db	1148	GAGAAAGGTTTGTGTCAGAAATTCAGGTGGCTTTTCTGTGATGTCTGACACCTTCATGTTT	1207
OY	1021	ATAAATGCTCAGCAAGGGTTTATCTTCTCAGATGCCAAATATAGTAAATAAAAAA	1080
Db	1208	GTGAATGCTCGTGCCCGAGTTTACGAGTTCAGATGTCTCAAACTCAAAAAAGGCCAAA	1267
OY	1081	ATATCTGAAAAAATGAAATTTAAGAGGGGAGAAACAAAAAAGCAACTGCTCTAGAA	1140

Db 1268 ACATCAGAAAGACGTCGAGTCAGAGTTCAAGAAACAAAAAGAACTGGTTCAGAG 1327
 Oy 1141 AACCAACCCAAATGGGAGGACCTGACCTGAGATTTAAAGAAATTTAGGACAGAAATAG 1200
 Db 1328 ACCAATCCAAATGGGAGGCGGTGACCTGACCTGAAAGAAATGCAACCAAAATAG 1387
 Oy 1201 GAGAGTGAAGCTCTGTTGGTCCAGTCAAGTACTGATTTGCAAGTATGACCGAACA 1260
 Db 1388 GAGACGAAAGCCCTTGTTGGTCCAGGCGGCTGCTGATTTGTCAGCATGACCGCAG 1447
 Oy 1261 TGTTCACAGCTGAGAGACTATATCACTCTTGGAGCGGAGCCCTTATTTAGAGCTTAC 1320
 Db 1448 TGTCTCCAGCTGAGAGACTACTGTCCTCCGCGGAGCAAAACCTTCTGCTGGACTTAT 1507
 Oy 1321 AGGAAACCTTTGAGAGATAGCAAAAGCTGAAAGAGTGTGATGAATTTAGGAAGAA 1380
 Db 1508 AGGAAACCTTTGAGAGATAGCAAAAGCTGAAAGAGTGTGATGAATTTAGGAAGAGG 1567
 Oy 1381 GACAGTTCAAGAGATTAAGAAATCTCAAAAGACCTAAAGACCCCAAAACAAAGAA 1440
 Db 1568 GAGGCGCCGAAAGAGAACTCAAAATCTGCAAAAGACCAAAAGCCCTCCAAACAAAGAG 1627
 Oy 1441 CGGCTTCTACCAAGAAAGAAC---CTCAAAAGAAAAAGGAGAGTTGACCTTACT 1497
 Db 1628 CGTCTCCGCGCAAAAGGAGGCGCCCTCTCAAAAGGAGAAAGCAAGAGCTGACTGACG 1687
 Oy 1498 CAATAGTGAAGAAACCTGAAGAACTGGAAGAGAGAGTGTGAGAGAGATATCT 1557
 Db 1688 CAAGTTTGGGTCGCTGAGAGAGCGCGGAGAGCAAGGCTTTGAGAGAAACCTGTGC 1747
 Oy 1558 CGAGAAATTAAGCAGTAGCCAGAAAGCTGCGGAGAAATTAAGCATGAAGAAATTTGAT 1617
 Db 1748 AGACAGACCAAGTAGAGCCAGAAAGCTGCGGAGTAGAGATTAAGGTGAATCAATTTGAT 1807
 Oy 1618 GTAATTTGTCTCGGATGCTCTTTCGAAATCTGAAAGAACCCCTCACTATCATCAT 1677
 Db 1808 TTAATTTGTCTCGGATGCTCTTTCGAAATCTGAAAGAACCCCTCACTATCATCAT 1867
 Oy 1678 CCGCTTCTGAGTGGAGGAGCCCTATGCTGACAAAGGTTCTCATGATGAAGAGCA 1737
 Db 1868 CCGCTTCTGAGTGGAGGAGCCCTATGCTGACAAAGGTTCTCATGATGAAGAGCA 1927
 Oy 1738 AGATAGTGTCTCTTATGAGCAGAGCTAACCTTTGTTGCGAGCTTGAATTTACAG 1797
 Db 1928 AGATAGTGTCTCTTATGAGCAGAGCTAACCTTTGTTGCGAGCTTGAATTTACAG 1987
 Oy 1798 GCGATAGGCTTGGGAAACCTCTGAGGTTTACTTTTATATACGAGGTTCACTGAG 1857
 Db 1988 GCTAGCAGGCTTGGGAAACCTCTGAGGTTTACTTTTATATACGAGGTTCACTGAG 2047
 Oy 1858 GAACAGGCTATCTACTGCTTTGCGGAAAGAAAGGCTTTTGAAGAACTCATTAAG 1917
 Db 2048 GAGCAGGCTATCTACTGCTTTGCGGAAAGAAAGGCTTTTGAAGAACTCATTAAG 2107
 Oy 1918 GAAAAAGCAAGCATGTTGTCCCTGAAGAAAGAGAGCAGATGAACCAAACTTAGAC 1977
 Db 2108 GAAAAAGCAGCATGTTGTCCCTGAAGAAAGAGAGAGATGAACCAAACTTAGAC 2167
 Oy 1978 CTAAGTAAGAGCAGCATCTGCAAGATGTTTCCACTGACACTGGAAGAGCCGCTGAG 2037
 Db 2168 CTAGCAAGAGGCTCAGAGCCCTGATGCTCCACGAGCACTCGGAAAGCCGCTGAG 2227
 Oy 2038 GAACAGATGTTGATCAGCAAGAAAGTGTGATGCGGATGCGGATTTGGAAGTACGCT 2097
 Db 2228 GAGCAGATGTTGATCAGCAAGAAAGTGTGATGCGGATGCGGATTTGGAAGTACGCT 2287
 Oy 2098 CCATCTCTGATCAGCTGCTGAGGAGTGTGATGATGAACCCGCTGACTTTAGAGTTGAGAT 2157
 Db 2288 CCATCTCTGATCAGCTGCTGAGGAGTGTGATGATGAACCCGCTGACTTTGAGAGTTGAT 2347
 Oy 2158 TACATCTCTACTCCAGAAATGTGCTGAGCGCAAGAGTATCATGATTTAATTCGCTCT 2217

Db 2348 TACATCTCTACTCCAGAACCTGAGCTGTGTGAGCGCAAAAGTGTGAGGAGCTCATTTGCTCG 2407
 Oy 2218 TTAATTAACGGCCGCTCTACAGCAGTGCATCTCCGCTACTTACAAAGCTCCC 2277
 Db 2408 CTGACAGTGGCGGCTCTTACAGCAGCTGCTAGCATGCTGCTACTTACGAGGAGCTC 2467
 Oy 2278 GTGCTCTGATGAGTTTGAACCTTACAGGAGCTTTCTCTCACTTCCGAGGAGCTTG 2337
 Db 2468 GTGCTCTGATGAGTTTGAACCTTACAGGAGCTTTCTCTCACTTCCGAGGAGCTTG 2527
 Oy 2338 TTTACAGAGATCTCCAGCAATGATTAAGTTCACAACTCACTTCTTCACTTCACTTC 2397
 Db 2528 TTTACAGAGATGATGAGCAGCAGCTGATGAGCTCCAACTTACGCTCTTCACTTCACTTC 2587
 Oy 2398 CCCAGACTACGATTTCTGTTGCTCCCTCTCTCTCATGCAAGGCGAGTTGTTGAGAG 2457
 Db 2588 CCGCTTTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2647
 Oy 2458 CTGAAACAAAGCAGCAGCTGATGCGGAGCAGCAGCTGCGCAATTTACAGAGATCT 2517
 Db 2648 CTGAAACAAAGCAGCAGCTGATGCGGAGCAGCAGCTGCGCAATTTACCGCAGACTCA 2707
 Oy 2518 GAAACCTTTCCGAGTCAAGAGATTAATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2577
 Db 2708 GAGACACTCTGAGTCAAGAGATTAATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2767
 Oy 2578 CCAGGCTGAATCCCAAAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2637
 Db 2768 CCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2827
 Oy 2638 TTAGCAGCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2697
 Db 2828 CTGCGCAGCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2887
 Oy 2698 CAGCTTTATGATTTTCACTTACACCTCTTTTCAAGAGTGTATCAAAAGGAGGAGA 2757
 Db 2888 CAGCTGACAGACTCTGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2947
 Oy 2758 AAGTGA 2763
 Db 2948 AAGTGA 2953

RESULT 8
 BC026792
 LOCUS
 DEFINITION
 Mus musculus, excision repair cross-complementing rodent repair deficiency, complementation group 4, clone MGC:30522 IMAGE:4911382, mRNA, complete cds.
 ACCESSION
 BC026792.1 GI:20073214
 VERSION
 BC026792.1
 KEYWORDS
 MGC.
 SOURCE
 house mouse.
 ORGANISM
 Mus musculus.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sclerognathi; Muridae; Murinae; Mus.
 REFERENCE
 1 (bases 1 to 3235)
 AUTHORS
 Strausberg, R.
 TITLE
 Direct Submission
 JOURNAL
 Submitted (04-APR-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
 REMARK
 NIH-MGC Project URL: <http://mgc.nci.nih.gov>
 Contact: MGC help desk
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: Jeffrey E. Green, M.D.
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Baylor College of Medicine Human Genome Sequencing Center
 Center code: BCM-HGSC
 Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>

Contact: amg@bcm.tmc.edu
 Gunaratne, P.H., Garcia, A.M., Lu, X., Huiyk, S.W., Hale, S.M.,
 Yoon, V.S., Kowis, C.R., Lawrence, S., Martin, R.G., Muzny, D.M.,
 Richards, S., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found
 through the I.M.A.G.E. Consortium/URL at: <http://image.llnl.gov>
 Series: IRK Plate: 41 Row: j Column: 17.

FEATURES

source

Location/Qualifiers
 1..3235

/organism="Mus musculus"
 /db_xref="taxon:10090"
 /map="FVB/N"
 /clone="MGC:30522 IMAGE:4911382"
 /tissue.type="Salivary gland, 10 week old female mouse"
 /clone_lib="MCL_CGAP_Sc2"
 /lab_host="DH10B"
 /note="Vector: PCMV-SPOrt6"
 24..2744

CDS

/codon_start=1
 /product="excision repair cross-complementing rodent
 repair deficiency, complementation group 4"
 /protein_id="AAH26792.1"
 /db_xref="GI:20073215"
 /db_xref="LocusID:50505"
 /translation="MAPLEVEROQVLELLDSDGLVVCARGLGTDRLLVHLRHCHP
 ACIVLINTOPAEFEYFINDKTEGVHILPRVTNLASREYVTOGIIPTSRIL
 LYVDPLITGRIPSLITGLIYVRAHRIIESQEFILRLRQKRRFIRATPDNAYAF
 DGFCHVERVNRNLEVRKLYMPREHVAANSLEQKREVEVLIHVMTPMLLITAI
 LDILNACLEKLNPSLEVEDLENAAGKPKDKIRHLDLMQLQALASIVD
 KLILRTLLQYLSQDVTFLNLESLERATEKVGQSGMLFLDASTMEFNARAYR
 LVPDKLNKKAITESEKSPSEVQTEKLELVESKWEALFDVKELEAEKSEALG
 PGHVLICASDDRTCOLRDYLSAGAEFLILRLYRKPEKQKEEVVNVNRKDGKR
 TTKSDRPRKAPRKERASAKRQPLKRRKQELIYVLSGSEPRPDRLKALEDCQT
 SSSPECGVEIKESFDLVNSDAIGILKEPLTIHPLUGSDPYALRLVLEVER
 YVYLVAELFVQLEIYRASRQPLRVFLIYGSTEQRYLTLRKEKAEFLI
 REKASVVEEREDNETDLARSGALDAPDTRKAGQENQSSIVDMREFR
 SELPSILHRGIDIEPVLEVDYILTELCYRKSVSLDLSLHSGRLYSOLASR
 YVRPVLIEFDPSKPSLAPRGAFQEMSSSVSKLTLTLHPRRLILMGPSHA
 TAELELKONKPODPAATAMATTADSELTIPESDRYRNPQDPLKMPVNAKNSL
 MNOVKRIAEATLSIERLTTIIGHSNKAQLHDFLHTATVADLVSKRVRK"

BASE COUNT 829 a 838 c 859 g 709 t
 ORIGIN

Query Match 48.5%; Score 1872; DB 10; Length 3235;
 Best Local Similarity 80.4%; Pred. No. 0;

Matches 2206; Conservative 0; Mismatches 535; Indels 3; Gaps 1;

QY 23 GGCAGCCGGCTCGACGAGATTGCCATGGCCGCTGCTGAGTAGACGACGAGCTGTGC 82
 Db 1 GCGTGTGGGGCGAAGCGGAGATGTGGCGCGCTGTGAGTAGACGAGCAGAGTGC 60
 QY 83 TGGAACTGCTGCAGACGAGGGGCTAGTAGTGGCGCGCGGGCTGGCGGCGGAGCGGC 142
 Db 61 TGGAACTGCTGCAGACGAGGGGCTGTGTGTGTGTGCGCGGAGCTGGCGACCGAGCGGC 120
 QY 143 TCCCTACCACTTCTTCAGCTGCATGCGACCGAGCTGCTGCTGTGCTGCTACAA 202
 Db 121 TCCCTACCACTTCTTCAGCTGCATGCGACCGAGCTGCTGCTGTGCTGCTACAA 180
 QY 203 CGCAGCCGGCGGAGGAGAGTATTTATCAATCAGCTGAGAGATGAGAGAGTGAACACC 262
 Db 181 CGCAGCCGGCGGAGGAGAGTATTTATCAATCAGTGAAGATGGAAGAGAGTGAACATC 240
 QY 263 TCCCTGCGCGGTATCAAAATGAAATCAAGCAAGTGGTATGAGTTTACACAGAG 322
 Db 241 TCCCTGCGAGGTGCAAAATGAAATGCAAGTACAGTGGTATGAGTACACAGAG 300
 QY 323 GTGGTGTATATTTGCGACAGTAGATACTGTGTGTGACTTGTGACTGATGAGATAC 382
 Db 301 GTGGTATATATTTGCAACAAGCGAGATACGTGTGTGATTTCTTGACAGGTGAGATAC 360
 QY 383 CTTCAGATTTAATTACTGGCATCTTGGTGTATVAGAGCCACAGAAATATGAGTCTTGTCT 442

Db 361 CTTCAGATTTAATTAACTGGCATCTTGGTGTATVAGAGCCACAGAAATATGAGTCTTGTCT 420
 QY 443 AAGAGCATTCATCTTGGCGCTCTTTGGCCAGAGAAAAACAAACGGTGTATTTAATAACTT 502
 Db 421 AGGAGCATTCATCTTGGCGCTCTTTGGCCAGAGAAAAACAAACGGTGTATTTAATAACTT 480
 QY 503 TCACAGACATTCATCTTGGCGCTTGTATGATGCTGTTTGTGATGCTGAGAGAGTGTAGAA 562
 Db 481 TCACAGACATTCATCTTGGCGCTTGTATGATGCTGTTTGTGATGCTGAGAGAGTGTAGAA 540
 QY 563 ATCTTTTGTGAGAGAACTGATCTGTGGCCAGAGTTCATGATGACAGTAACTCATTTT 622
 Db 541 ACCTTTTGTGAGAGAACTGATCTGTGGCCAGAGTTCATGATGACAGTAACTCATTTT 600
 QY 623 TAGACAGCAAAACCTGAGAGTGTATAAATCCATGTTTGTATGACACCTACCTGCTTG 682
 Db 601 TGGAAACAACACACCTGAGAGTGTATGAGATTCACGTCGATGACATCCGCGCATGCTTG 660
 QY 683 CTATACAGACATCTGATCTGATCTGATCTTAAATGATGATCTTAAGAGAACTAAATGCCATA 742
 Db 661 CCATTCAGAGCGCATCTGATCTGATCTTAAATGATGATCTTGAAGAACTGAATGTCACA 720
 QY 743 ACCCATGCTTGAAGTGAAGATTTATCTTTAGAAAACTGATTTGCAAAACCTTTTGACA 802
 Db 721 ACCGTCACATGAGTGAAGATTTGTCACATGAAAACTGCTTGGCAAGCATTTGACA 780
 QY 803 AGCAATCCGCGCATCTGATCTGATCTTGTGGCCACAGCTGGAGCCGACAGCTAAATCCT 862
 Db 781 AGCAATCCGCGCATCTGATCTGATCTTGTGGCCACAGCTGGAGCCGACAGCTAAATCCT 840
 QY 863 TAGTTCAGATTTGAAGATTTAGCAACTTGTGCGAGTATCTGTCAGATGATTTGTG 922
 Db 841 TGGTCAGAGATTTAAATATCTAGAAACCTTACGCACTATCTCTCAGATATATTTGTG 900
 QY 923 TCACATTTCTTAATCTTCTGGAATCTGTGAGAGACGAGAAAAACCTTTTGTGCAATTT 982
 Db 901 TTACATTTCTGAATCTTCTGGAATCTGTGAGAGACGAGAAAAACCTTTTGTGCAATTT 960
 QY 983 CAGGTGGCTGTTCTGTGACCTCCGACACCTGATTTAATAATGTCGAGCAAGGGTTT 1042
 Db 961 CAGGTGGCTGTTCTGTGAGCTCCGACCTCAATGTTTGTGAATGTCGAGCAAGGGTTT 1020
 QY 1043 ATCATCTTCCAGATGCGCAAAATGAGTAAAGAAAAAATATCTGAAAAAATGAGAAATTA 1102
 Db 1021 ACCGATGTCAGATGTCGCAAAATGAGTAAAGAAAAAATATCTGAAAAAATGAGAAATTA 1080
 QY 1103 AAGAGGGAAGAAACAAAGAAAGAACTGCTCTAGAAAGCAACCCAAAGTGGAGAGCAC 1162
 Db 1081 CAGAAAGTTCAGAAACAAAGAAAGAACTGCTCTAGAAAGCAACCCAAAGTGGAGAGCAC 1140
 QY 1163 TGACTGAGATTTAAAGAAATTAGAGCAAGAAATAGAGAGAGTGAAGTCTTGTGTGTC 1222
 Db 1141 TGACTGAGATTTAAAGAAATTAGAGCAAGAAATAGAGAGAGTGAAGTCTTGTGTGTC 1200
 QY 1223 CAGGTCAAGTACGATTTGTGCAAGTATGACGAGCAAGATGTTCCCACTGAGACATATA 1282
 Db 1201 CAGGTCAAGTACGATTTGTGCAAGTATGACGAGCAAGATGTTCCCACTGAGACATATA 1260
 QY 1283 TCACCTTTGAGAGCGGAGGCTTCTTATTTAGAGCTCTACAGAAACCTTTGAGAGAGATA 1342
 Db 1261 TGTCCGCGGAGGAGAACTTCTGCTGCGACTTTATAGGAAACCTTTTGAAGAGATG 1320
 QY 1343 GCAAAAGTGAAGAGTGTGATGTAATTTAGAGAGAAAGCAAGTTCAAAGAGATTTAGAA 1402
 Db 1321 GCAAAAGTGAAGAGTGTGATGTAATTTAGAGAGAAAGCAAGTTCAAAGAGATTTAGAA 1380
 QY 1403 AATCTCAAAAAGACCTAAAGACCCCAAAAAGAAAGAGGCTTTTACCAAGAAAGAA 1462
 Db 1381 AATCTCAAAAAGACCTAAAGACCCCAAAAAGAAAGAGGCTTTTACCAAGAAAGAA 1440
 QY 1463 CC---CTCAAAAAGAAAGAGGAGTGTGACCTTAACCTCAAAATGCTAGGAAACCTGAG 1519
 Db 1440 CC---CTCAAAAAGAAAGAGGAGTGTGACCTTAACCTCAAAATGCTAGGAAACCTGAG 1519

D	b	1441	CCCTCTCAAAAGAGAAAGAAAGCAAGACGCTGACCTGACCGAAGTTTGTGGGGTCCGCTAGG	1500
Q	y	1520	AACTGGAAGAGGAAGAGATGTGAGGAAGATATCTGCGAAGATTAAGACTAGCCAG	1579
D	b	1501	AGCGCGCGGAGCAAGAGCTTTGGAGGAAGACCTGTGCGAGACAGCCAGTGTAGCCAG	1560
Q	y	1580	AAAGCTGCCCGGAAGAAATTAAAGCATGAAGATTTGGATTAATTGTGATCGGATGCTG	1639
D	b	1561	AAGGCTGCGGGGTAGAGATTTAAAGCTGAATCATTTTGATTTAAATGTGTCTGATGGG	1620
Q	y	1640	CTTTGGAACTCTGAAAGAAGCCCTCTACTATCATCTATCCGCTTCGGGTTGACGACAGC	1699
D	b	1621	CCATATGGCATCTGAAAGAAGCCCTGACCATCATCCATTCGCTTCAGGTGCGAGCGACC	1680
Q	y	1700	CCATATGCTTGACAAAGGATCTACATGAATGGAGCCAAAGATAGCTGGTCTTTATGACG	1759
D	b	1681	CCATATGCCCTGAGCGGGTACTGACAGAGGTGGAGCCCAAGATACCTGGTCTGATATGTG	1740
Q	y	1760	CAGAGCTAACTTTGTTCCGCGACGCTTGAATTTTACAGGGCCGTTAGCGCTGGGAACCTC	1819
D	b	1741	CAGAGCTACCTTTGTTTGGCAACTGGAATCTACAGGGCTTACAGAGGCTTGGAAACCTC	1800
Q	y	1820	TGAGGGTTTACTTTCTTATATACGAGAGTTTCAACTAGAGAACAAAGCTATCTCACTGCT	1879
D	b	1801	TGAGGGTTTACTTTCTTATATACGAGAGGCTCCACCGAAGAGAGCGCTACTGACTGCTC	1860
Q	y	1880	TGCGGAAAGAAAGAGAACTTTGAAAAAATCTATAAGGAAAAAGCAAGATGGTTGTC	1939
D	b	1861	TGCGGAAAGAAAGAGAAAGCTTTGAAAAAATCTATAAGAAAAAAGCTAGCTGTTGTC	1920
Q	y	1940	CTGAAGAAAGAGAGCAGAGATGAAGAAACAATTTAGACTAGTAAAGAGGCACAGATCTG	1999
D	b	1921	CCGAGAGAAAGGAGAGCAGAGATGAACCAACCTGAGCTTACCAAGAGGCTACAGAGCC	1980
Q	y	2000	CAGATGTTTCCATGACACTCGGAAAGCCGCTGGCAGAGAACAGAAATGTACACAGCAA	2059
D	b	1981	TGAGTCTGCCAGGAGACACTCGSAAAGCCGGTGGCCAGAGACAGAAATGGACCAAGTCA	2040
Q	y	2060	GCATAGTTGTGATATGCTGAAATTTGAAATGAGAGCTTCATCTCTGATCCATGCTGGG	2119
D	b	2041	GCATCGGTGTGGACATATGCGGATTTTGGAGGAGACCTCCATCTCTGATCCATCGGGG	2100
Q	y	2120	GCATTTGACATTTGAACCCGTGACTTTAGAGTTGGAGATTTACATCTCTACACAAATGT	2179
D	b	2101	GTAATGACATAGAGCCAGATGACACTGTGAGGTGGGTGATTTACATCTCTGACACCTGAGCT	2160
Q	y	2180	GGGTGGAGCCCAAGATATGATACGATTTTAATGCGCTTTAAATAAAGCCGCGCTTCA	2239
D	b	2161	GTTGGAGGCCCAAAAGTGTGAGGACCTATATTTGGCTCGCTGCACAGTGGCGCGCTTACA	2220
Q	y	2240	GCCAGTGCATCTCATATGTCGCCGTACTATACAAAGCGTCCCGCTTCTGATTAAGTTTGAAC	2299
D	b	2221	GCCAGTGCATAGCAGATGTCTCGTACTATACCGGAGGGCTGTGCTGCTATTTAGTTTGAAC	2280
Q	y	2300	CTAGCAAGCTTTCTCTCTCACTTCCGAGGTGCTTGTTTAGAGAGATCTTCAGCAATG	2359
D	b	2281	CGAGCAAGCCCTTCTCTCTGCGCCGACGAGGTGCTTTCTTCAGAGAGATGAGCAGACGC	2340
Q	y	2360	ACATTAATTTCAAAATCACTCTTTTAACTTCACTTCCCGAGACTACAGGATCTGCTGCT	2419
D	b	2341	ATGTAGACTCAAAAGCTTACGCTCTCTCACTTCACTTCCCGGCTTTGGCGCTCTGTGCT	2400
Q	y	2420	GCCGCTCTCTCAATGCAAGCGCGAGATTTTGGAGAGCTGAACAAAGCAAGCCACAGC	2479
D	b	2401	GCCGCTCACCCCACAGCCACAGCCGAGACTGTTTGAAGAGCTGAAGCAAGAACACAGC	2460
Q	y	2480	CTGATGCGGCGAAGACACTGTGCCATTTACAGCAGATTTCTGAACCTTCCGAGTACAGA	2539
D	b	2461	CTGATGCGGCGCAAGCATATGGCCATCAACGCGAGACTCAAGACATCACTGCTAGTGGACA	2520
Q	y	2540	AGTAAATCCGCGCCCAAGACTTCTTTTAAATAAGCAGAGGGTGAATGCCAAAAACT	2599
D	b	2521	GATTAATCCCGCGCCCAAGACTTTTGTCTGAAGATGTCTGGGTCAACGCCAAGAACT	2580

QY	2600	GCCTCTCTTATGACACGCTTAAAGCATTCGAGATTTGACAGCCCTGTCCACAAAGC	2659
Db	2581	GTCCGTCCCTTATGACCAACAGTGAAGAACATTCGAGAGCTGCCACCTGTCCCTGGAAC	2640
QY	2660	AGCTCAGCAGTATTCCTGGGGAATGCTGCMAATGCCAAGACCTTATCATTTCAATTCACA	2719
Db	2641	GACTCACAACCATTCCTGGGACATCGGGGAATATGCCAACACACTGCACACTTCTCTGCACA	2700
QY	2720	CCTCTTTTGCAAGTCTATCAAAAGGAAGAAAGGAAAGTGA	2763
Db	2701	CTGCCTATCGACACTTGGTGTCCAAAGGACAGATGAGAAATGA	2744
RESULT 9			
LOCUS	AX332256	35641 bp	DNA
DEFINITION	Sequence 2765 from Patent WO0194629.	Linear	PAT 09-JAN-2002
ACCESSION	AX332256		
VERSION	AX332256.1	GI:18122890	
KEYWORDS	human.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Euhayloia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
AUTHORS	1 Young, P.E., Augustus, M., Carter, K.C., Ebner, R., Endress, G., Horrijan, S., Soppet, D.R. and Weaver, Z.		
TITLE	Cancer gene determination and therapeutic screening using signature gene sets		
JOURNAL	Patent: WO 0194629-A 2765 13-DEC-2001;		
FEATURES	Avalon Pharmaceuticals (US)		
SOURCE	Location/Qualifiers		
	1..35641		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
BASE COUNT	10575 a 7169 c 7323 g 10574 t		
ORIGIN			
Query Match	47.4%:	Score 1829.4:	DB 6: Length 35641;
Best Local Similarity	99.4%:	Pred. No. 0:	
Matches 1836:	Conservative 0:	Mismatches 11:	Indels 0: Gaps 0:
QY	2011	ACTGACACTCGGAAAGCCGCTGGCCAGAGACAGATGTGACACAGCAAGCATAGTTGTG	2070
Db	29522	ACTTACTTTTCTCTCTGTAGTGTGGCCAGGACAGATGTGTACACAGCAAGCATAGTTGTG	29581
QY	2071	GATATGCGTGAATTTTCGAAGTGAAGCTTCATCTGTATCATTCGTGGGCGATTGACATT	2130
Db	29582	GATATGCGTGAATTTTCGAAGTGAAGCTTCATCTGTATCATTCGTGGGCGATTGACATT	29641
QY	2131	GAACCCGAGACTTTGAGAGTTGGAGATTAACATCCATCCATCCAGAAATGGGTGGAGCGC	2190
Db	29642	GAACCCGAGACTTTGAGAGTTGGAGATTAACATCCATCCATCCAGAAATGGGTGGAGCGC	29701
QY	2191	AAGATATCAGTGAATTAATGAGTCTTTAATAAAGCGCCCTCTACAGCCAGTGCATC	2250
Db	29702	AAGATATCAGTGAATTAATGAGTCTTTAATAAAGCGCCCTCTACAGCCAGTGCATC	29761
QY	2251	TCCATGTCCTGCTACTACAGAGGCTCCGCTCTGTGATTAAGTTGACCTTACAGAACGCT	2310
Db	29762	TCCATGTCCTGCTACTACAGAGGCTCCGCTCTGTGATTAAGTTGACCTTACAGAACGCT	29821
QY	2311	TTCTCTGTCACGTCCTCCGAGGTGCTTTGTTAAGAGGATCTCCACCAATGACATTAGTTC	2370
Db	29822	TTCTCTGTCACGTCCTCCGAGGTGCTTTGTTAAGAGGATCTCCACCAATGACATTAGTTC	29881
QY	2371	AAACTCAGCTTTCTTACACTTCATCTCCCAAGACATACAGGATTTCTGTGCCCCCTTCTCT	2430
Db	29882	AAACTCAGCTTTCTTACACTTCATCTCCCAAGACATACAGGATTTCTGTGCCCCCTTCTCT	29941
QY	2431	CATGCAACGGGAGTGTGTTGAGAGCTGAACAACCAAGCCACAGCTGATGCGCG	2490

Db 29942 CATGCAACGGCGAGTGTGTTGAGAGCTGAACAACGACCAAGCCTGATGCGGGC 30001
QY 2491 ACAGCAGTGGCATTACACAGATTCTGAACCCCTCCGAGTCAAGAGATATATCCT 2550
Db 30002 ACAGCAGTGGCATTACACAGATTCTGAACCCCTCCGAGTCAAGAGATATATCCT 30061
QY 2551 GGTCCCAAGACCTTCTGTTAAATAATGCCAGGGGTGAATGCCAAAACGCCGCTCTTG 2610
Db 30062 GGTCCCAAGACCTTCTGTTAAATAATGCCAGGGGTGAATGCCAAAACGCCGCTCTTG 30121
QY 2611 ATGCACACAGTTAAGACATGCCAGATTAGCAGCCCTCTCAAGACAGCTCACAGT 2670
Db 30122 ATGCACACAGTTAAGACATGCCAGATTAGCAGCCCTCTCAAGACAGCTCACAGT 30181
QY 2671 ATTCTGGGGAATGCTGCAAAATGCCAAGACGTTTATGATTATTCACACCTCTTTGCA 2730
Db 30182 ATTCTGGGGAATGCTGCAAAATGCCAAGACGTTTATGATTATTCACACCTCTTTGCA 30241
QY 2731 GAAGTCGTATCAAAAGAAAAGGAAAAGTGAACAGTGGCTGTTTCTTATCCAT 2790
Db 30242 GAAGTCGTATCAAAAGAAAAGGAAAAGTGAACAGTGGCTGTTTCTTATCCAT 30301
QY 2791 GCGTGTACTTTACAGCGGCTCTTCGACACATCATAGTCAATATATATATGTTTG 2850
Db 30302 GCGTGTACTTTACAGCGGCTCTTCGACACATCATAGTCAATATATATATGTTTG 30361
QY 2851 CTATTTTCATCTTTTCCAAATGCTCTTAATGATTGATGATGACGACGAGAGATTC 2910
Db 30362 CTATTTTCATCTTTTCCAAATGCTCTTAATGATTGATGATGACGACGAGATTC 30421
QY 2911 TCTGTGAACCTGCAAGTTAGCATCACTGAACCTGCTGCTGCTCTTTTCTCTCC 2970
Db 30422 TCTGTGAACCTGCAAGTTAGCATCACTGAACCTGCTGCTGCTCTTTTCTCTCC 30481
QY 2971 TGCACCGTCTATGCGGGGCTTGACATGTTCTTTTAAATGAGGTTTGCAGATCAGT 3030
Db 30482 TGCACCGTCTATGCGGGGCTTGACATGTTCTTTTAAATGAGGTTTGCAGATCAGT 30541
QY 3031 AAGTTTCTTACAGTATGATACAGAAAGTGAAGCTTATACCTGATCTTAACATCTCAT 3090
Db 30542 AAGTTTCTTACAGTATGATACAGAAAGTGAAGCTTATACCTGATCTTAACATCTCAT 30601
QY 3091 TGAAGAAGATATGCTAGCCTGCGATGAGCGGTGACGAGGAGGAAAAGACAGCACAA 3150
Db 30602 TGAAGAAGATATGCTAGCCTGCGATGAGCGGTGACGAGGAGGAAAAGACAGCACAA 30661
QY 3151 GAAAGCTACATTTTAAACAGTCTGTTATCTAGGCAACATTAATACAGCTTAAT 3210
Db 30662 GAAAGCTACATTTTAAACAGTCTGTTATCTAGGCAACATTAATACAGCTTAAT 30721
QY 3211 GCATTTATACCATGTCCTGCTGCTCCAAATCTGCTCTTGTGCTGTTGCTGCTGGA 3270
Db 30722 GCATTTATACCATGTCCTGCTGCTCCAAATCTGCTCTTGTGCTGTTGCTGCTGGA 30781
QY 3271 GCGTTGAAGTATGTTTGTGTGAAGAAATCATGTTGACCTTTGTCTACAAAGAGCT 3330
Db 30782 GCGTTGAAGTATGTTTGTGTGAAGAAATCATGTTGACCTTTGTCTACAAAGAGCT 30841
QY 3331 TCTGGAACCTGAGAAGAAACATCTTTGGCATCTGACAGAGTCTCTACACAT 3390
Db 30842 TCTGGAACCTGAGAAGAAACATCTTTGGCATCTGACAGAGTCTCTCTACACAT 30901
QY 3391 TTTCTTACAGTCAATCTTGTGCTGCTCTGAGGAAATTTCAATGAGGCTTCTTACT 3450
Db 30902 TTTCTTACAGTCAATCTTGTGCTGCTCTGAGGAAATTTCAATGAGGCTTCTTACT 30961
QY 3451 ACTAATTCAGACAGTCTCTCAAAAACGTTGACTAGTCTTCTAATGACCTTAACATA 3510
Db 30962 ACTAATTCAGACAGTCTCTCAAAAACGTTGACTAGTCTTCTAATGACCTTAACATA 31021
QY 3511 TGTACAT 3570
Db 31022 TGTACAT 31081

QY 3571 CTGTTGCAAAAGTTAATGATGAGAGGCTTAGAATTCATATTTTGCATATATCA 3630
Db 31082 CTGTTGCAAAAGTTAATGATGAGAGGCTTAGAATTCATATTTTGCATATATCA 31141
QY 3631 GTCTCTAATATACAGATCCCTAAGTCCAGCTGGCTAGTACAGAGTTTTCAGACT 3690
Db 31142 GTCTCTAATATACAGATCCCTAAGTCCAGCTGGCTAGTACAGAGTTTTCAGACT 31201
QY 3691 CCTGCTTTCAGCTCTTATATCTTAAGACACAGCATATATCTTGAATAACAC 3750
Db 31202 CCTGCTTTCAGCTCTTATATCTTAAGACACAGCATATATCTTGAATAACAC 31261
QY 3751 TAATTGACAGTACGAGAGATGACACACCTGACCCCTGCGGCGAGAGTGAAGT 3810
Db 31262 TAATTGACAGTACGAGAGATGACACACCTGACCCCTGCGGCGAGAGTGAAGT 31321
QY 3811 TTGCTCTATTTACAAAAAGAAAAGAAATACAACTTAAGCTCA 3857
Db 31322 TTGCTCTATTTACAAAAAGAAAAGAAATACAACTTAAGCTCA 31368
RESULT 10
AA409578 35641 bp DNA linear PAT 14-JUN-2002
LOCUS
DEFINITION Sequence 2225 from Patent WO0229103.
ACCESSION AA409578
VERSION AA409578.1 GI:21442283
KEYWORDS
SOURCE
ORGANISM human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE
1
AUTHORS Alvarez, C., Horne, D., Perez-da-Silva, S. and Vockley, J.G.
Gene expression profiles in liver cancer
Patent: WO 0229103-A 2225 11-Apr-2002;
JOURNAL
FEATURES
location/Qualifiers
source 1..35641
/organism="Homo sapiens"
/db_xref="taxon:9606"
/note="EMBL/GenBank Accession No. U76568"
BASE COUNT 10575 a 7169 c 7323 g 10574 t
ORIGIN
Query Match 47.4%; Score 1829.4; DB 6; Length 35641;
Best Local Similarity 99.4%; Pred. No. 0;
Matches 1836; Conservative 0; Mismatches 11; Indels 0; Gaps 0;
QY 2011 ACTGACACTCGGAAAGCCGCGGCGCAGAAAGATGTACACACCAGATAGTTGTG 2070
Db 29522 ACTTACTTTTCTCTGTAGTGGCCAGAAAGATGTACACACCAGATAGTTGTG 29581
QY 2071 GATATGCGTAATTTGGAAGTACGCTTCATCTGATCATCTGCGGCAATTGACATT 2130
Db 29582 GATATGCGTAATTTGGAAGTACGCTTCATCTGATCATCTGCGGCAATTGACATT 29641
QY 2131 GAACCCGTAATTTAGAGTTGAGATTCATCTCCTCAGAAATCTGCTGAGCGCC 2190
Db 29642 GAACCCGTAATTTAGAGTTGAGATTCATCTCCTCAGAAATCTGCTGAGCGCC 29701
QY 2191 AAGAGTACAGATTAATTTGAGGCTTTAAATACAGGCGGCTCTACAGCAGTGATC 2250
Db 29702 AAGAGTACAGATTAATTTGAGGCTTTAAATACAGGCGGCTCTACAGCAGTGATC 29761
QY 2251 TCCATGTCGCGTACTACAAAGCCTCCGCTTCTGATGAGTTTGACCTAGCAAGCT 2310
Db 29762 TCCATGTCGCGTACTACAAAGCCTCCGCTTCTGATGAGTTTGACCTAGCAAGCT 29821
QY 2311 TTCTCTCTACCTCCGAGGCTCTTGTTCAGAGATCTCCAGCAATGATAGTTCC 2370
Db 29822 TTCTCTCTACCTCCGAGGCTCTTGTTCAGAGATCTCCAGCAATGATAGTTCC 29881

OY	2371	AACTCAGCTCTCTTACACTTCACCTTCCCCAGACACTACGAGATTCTGTGGGCCCTCTCT	2430
OY	29882	AAACGACCTCTCTTACACTTCTACTTCCCCACACACTACGAGATTCTGTGGGCCCTCTCT	29941
OY	2431	CATGCAACGGGGGAAATTTTGTAGAGAGCTGAACAAACAGCAAGCCACAGCTCATCGGG	2490
Db	29942	CATGCAACGGGGGAAATTTTGTAGAGAGCTGAACAAACAGCAAGCCACAGCTCATCGGG	30001
OY	2491	ACAGCACTGGCCATTACAGCAGATTCTGAACCCCTTCCGAGTCAGAGAATATAACTCT	2550
Db	30002	ACAGCACTGGCCATTACAGCAGATTCTGAACCCCTTCCGAGTCAGAGAATATAACTCT	30061
OY	2551	GGTCCCAAGACTTCTGTGTAAATATGCCAGGGGGGAATGCCAAAACATGCCGCTCTTG	2610
Db	30062	GGTCCCAAGACTTCTGTGTAAATATGCCAGGGGGGAATGCCAAAACATGCCGCTCTTG	30121
OY	2611	ATGCACCACTGTTAAGACATCGCAGAAATTAGCAGCCCTGTCAAGAGAGCTCACAGAT	2670
Db	30122	ATGCACCACTGTTAAGACATCGCAGAAATTAGCAGCCCTGTCAAGAGAGCTCACAGAT	30181
OY	2671	ATTCGGGGGAATGCTGCAAAATGCCAAACAGCTTATGATTTATTACACCTCTTTGCA	2730
Db	30182	ATTCGGGGGAATGCTGCAAAATGCCAAACAGCTTATGATTTATTACACCTCTTTGCA	30241
OY	2731	GAAGCTGATCAAAAAGAAAAAGGAAAAAGTATACAGTCATGGCTGTTTCTTATCCAT	2790
Db	30242	GAAGCTGATCAAAAAGAAAAAGGAAAAAGTATACAGTCATGGCTGTTTCTTATCCAT	30301
OY	2791	GCTCTACTTTTACGGGCTCCTTCCAGACATCATAGTCATTTAATTAATTAGGTTTG	2850
Db	30302	GCTCTACTTTTACGGGCTCCTTCCAGACATCATAGTCATTTAATTAATTAGGTTTG	30361
OY	2851	CTATTTCATTTCTTTTCCATGCTCTTAATGATTTAGGTGAGCCAGCAAGCCAGATTC	2910
Db	30362	CTATTTCATTTCTTTTCCATGCTCTTAATGATTTAGGTGAGCCAGCAAGCCAGATTC	30421
OY	2911	TCTCTGAACCTCTGCAGTTAGGATCACTTGAACCTTGCCTGCTCCTTCTTCTCTCC	2970
Db	30422	TCTCTGAACCTCTGCAGTTAGGATCACTTGAACCTTGCCTGCTCCTTCTTCTCTCC	30481
OY	2971	TGCACCGCTATAGCCGGGCTTAGCATGTTTCTTTTAAATAGAGTTTGTACAGATCAGT	3030
Db	30482	TGCACCGCTATAGCCGGGCTTAGCATGTTTCTTTTAAATAGAGTTTGTACAGATCAGT	30541
OY	3031	AAAGTTCTTACAAAGATTTACAGAGAGTGAAGAACTTACCTATCTTAACATCATCT	3090
Db	30542	AAAGTTCTTACAAAGATTTACAGAGAGTGAAGAACTTACCTATCTTAACATCATCT	30601
OY	3091	TAGAAAGCAATATGCTAAAGCTTGGCATGTGACGTGACGGAGGAAAAAGCAGGCACAA	3150
Db	30602	TAGAAAGCAATATGCTAAAGCTTGGCATGTGACGTGACGGAGGAAAAAGCAGGCACAA	30661
OY	3151	GAAACCTACCATTTTTTAAACAGAGCTTGTATATAGGCAACCTAATAATACATCTTAAT	3210
Db	30662	GAAACCTACCATTTTTTAAACAGAGCTTGTATATAGGCAACCTAATAATACATCTTAAT	30721
OY	3211	GCACCTATACCAGTATGCTGTGGCTCTCCAAATCTGTGTTTGTCTGTCTGTGA	3270
Db	30722	GCACCTATACCAGTATGCTGTGGCTCTCCAAATCTGTGTTTGTCTGTCTGTGA	30781
OY	3271	GCCTTGAACATGATGTTTGTGTAGGAAATCATGTCTGACCCCTTGTCTCAAAAGAGCCT	3330
Db	30782	GCCTTGAACATGATGTTTGTGTAGGAAATCATGTCTGACCCCTTGTCTCAAAAGAGCCT	30841
OY	3331	TCTGGAACACTGAGAAGAAACAATCTTTTGCATTCCTGACAGAGTTCTCTACACAT	3390
Db	30842	TCTGGAACACTGAGAAGAAACAATCTTTTGCATTCCTGACAGAGTTCTCTACACAT	30901
OY	3391	TTCTTCACGCTCATCTCTCTGCTGTCTGTCTTAAGAGAAATTTCAATGAGGCTTCTCTACT	3450
Db	30902	TTCTTCACGCTCATCTCTCTGCTGTCTGTCTTAAGAGAAATTTCAATGAGGCTTCTCTACT	30961

OY	3451	ACTAATTCGAAGCAGCTGCTCCCAAAACACTGGTGAGCTGTCTTTTAAAGACCCTTAACAATA	3510
Db	30962	ACTAATTCAAGAAGCATCTCTCCAAAACCTGGTGCAGTCTCTTAATGAGACCTTAACATA	31021
OY	3511	TGTACATATATCTATTAATTTCAATTTGTGCCAAATTTGATATTTTTAAAGCAAATGAATTAC	3570
Db	31022	TGTACATATATCTATTAATTTCAATTTGTGCCAAATTTGATATTTTTAAAGCAAATGAATTAC	31081
OY	3571	CTGTTTGGCAAAAGTTAATGATGAAAGAGCTCTTAGAATTTCTCAATTTTTTGCACATATTCA	3630
Db	31082	CTGTTTGGCAAAAGTTAATGATGAAAGAGCTCTTAGAATTTCTCAATTTTTTGCACATATTCA	31141
OY	3631	GCTCTCTAATATATCAGAGATCCCTTAAGTCCAGCTGGCTAGTTACAGAGTTTTTTAGACTT	3690
Db	31142	GCTCTCTAATATATCAGAGATCCCTTAAGTCCAGCTGGCTAGTTACAGAGTTTTTTAGACTT	31201
OY	3691	CCTGCTTTCTGAGCTTATATATCCCTTAAGCACACACATCATATCCTCTAGAAATACAAAC	3750
Db	31202	CCTGCTTTCTGAGCTTATATATCCCTTAAGCACACACATCATATCCTCTAGAAATACAAAC	31261
OY	3751	TAAATGGCAGTAGCGGAGATCGCACACACTGACCCCTGCTGGGCGACAGAGTAGACT	3810
Db	31262	TAAATGGCAGTAGCGGAGATCGCACACACTGACCCCTGCTGGGCGACAGAGTAGACT	31321
OY	3811	TTGTCTCTATTACAAAAAGAAAGAAATATACACTTAAGCTCA	3857
Db	31322	TTGTCTCTATTACAAAAAGAAAGAAATATACACTTAAGCTCA	31368
RESULT 11			
HUMERCC4G		35641 bp DNA linear PRI 22-MAR-1997	
LOCUS			
DEFINITION		Homo sapiens excision and cross link repair protein (ERCC4) gene,	
ACCESSION		complete genomic sequence.	
VERSION	L76568		
KEYWORDS	L76568.1 GI:1905925		
SOURCE		excision and cross link repair protein.	
ORGANISM		Homo sapiens (library: Human chromosome 16 cosmid library) DNA.	
REFERENCE			
AUTHORS		Homo sapiens	
TITLE		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
JOURNAL		Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	
MEDLINE		1 (bases 1 to 35641)	
PUBMED		Brookman,K.W., Lamerdin,J.E., Thelen,M.P., Hwang,M., Reardon,J.T.,	
REFERENCE		Sancar,A., Zhou,Z.Q., Walter,C.A., Parris,C.N. and Thompson,L.H.	
AUTHORS		ERCC4 (XPR) encodes a human nucleotide excision repair protein with	
TITLE		eukaryotic recombination homologs	
JOURNAL		Mol. Cell. Biol. 16 (11), 6553-6562 (1996)	
MEDLINE			
PUBMED		8887684	
REFERENCE		2 (bases 1 to 35641)	
AUTHORS		Lamerdin,J.E.	
TITLE		Direct Submission	
JOURNAL		Submitted (03-DEC-1996) J.E. Lamerdin, Human Genome Center,	
COMMENT		Lawrence Livermore National Laboratory, 7000 East Ave, Livermore,	
FEATURES		CA, USA, 94551 jane@agct.llnl.gov ow@tornak.llnl.gov	
SOURCE		GSDB:S:133826	
		Location/Qualifiers	
		1..35641	
		/organism="Homo sapiens"	
		/db_xref="taxon:9606"	
		/map="16p13.13"	
		/issue_1lb="Human chromosome 16 cosmid library"	
		2093..30272	
		/gene="ERCC4"	
		2093..2299	
		/gene="ERCC4"	
		/note="putative"	
		/db_xref="GDB:119113"	
		2300..3957	
		/gene="ERCC4"	
		/note="putative"	
		/number=1	
		/db_xref="GDB:119113"	

Dp	29702	AAGAGTACGATGATTTAATGCGGCTTTAAATTAACGGCGGCTCTACAGCCAGTGCAT	29761
Qy	2251	TCGATGTCCCGTACTACAGAGCTCCGCTGCTTCTGATTGAGTTGACCTTAGCAAGCT	2310
Dp	29762	TCCAGTCCCGGTACTACAGAGCTCCCGCTTCTGATTGAGTTGACCTTAGCAAGCT	29822
Qy	2311	TTCTCTACACTTCCCGAGGCTCGTTGTTACAGGAGATCTCCAGCAATACATTACTTCC	2370
Dp	29822	TTCTCTACACTTCCCGAGGCTCGTTGTTACAGGAGATCTCCAGCAATACATTACTTCC	29881
Qy	2371	AAACTCACTCTTTACACTTCTTCCACAGCTACGGATTTCTCGTGCCCTCTCGT	2430
Dp	29882	AAACTCACTCTTTACACTTCTTCCACAGCTACGGATTTCTCGTGCCCTCTCGT	29942
Qy	2431	CATGCAAGCGCGGAGTTGTTTAGGAGCTGAAACCAAGCAAGCCACGCTGATGCGGG	2490
Dp	29942	CATGCAAGCGCGGAGTTGTTTAGGAGCTGAAACCAAGCAAGCCACGCTGATGCGGG	30001
Qy	2491	ACAGCACTGGCCATTACAGCAATTCGTAACCCCTCCGAGTCAGAGAATTAATCCG	2550
Dp	30002	ACAGCACTGGCCATTACAGCAATTCGTAACCCCTCCGAGTCAGAGAATTAATCCG	30061
Qy	2551	GCTCCCAAGACTTCTTGTTAAAAATGCCAGGGTGATGCCAAAAATGCCGCTCTTG	2610
Dp	30062	GCTCCCAAGACTTCTTGTTAAAAATGCCAGGGTGATGCCAAAAATGCCGCTCTTG	30122
Qy	2611	ATGCACACAGTTAAAGAACTGCGAGATTAGAGCCCTGTCACAGAGCAGCTACAGAT	2670
Dp	30122	ATGCACACAGTTAAAGAACTGCGAGATTAGAGCCCTGTCACAGAGCAGCTACAGAT	30182
Qy	2671	ATTCGGGGAATGCTGCAAAATGCCAAACAGCTTATGATTTCATTCACACTTTTGA	2730
Dp	30182	ATTCGGGGAATGCTGCAAAATGCCAAACAGCTTATGATTTCATTCACACTTTTGA	30242
Qy	2731	GAAGTCGATCAAAAGGAAAAGGAAAAAGTAAACAGTATGGCTTTTCTTATCCCAT	2790
Dp	30242	GAAGTCGATCAAAAGGAAAAGGAAAAAGTAAACAGTATGGCTTTTCTTATCCCAT	30302
Qy	2791	GCCCTACTTTTACAGCGGCTCCCTCCAGACATCAATAGCTCATTTAATTAATGCTTG	2850
Dp	30302	GCCCTACTTTTACAGCGGCTCCCTCCAGACATCAATAGCTCATTTAATTAATGCTTG	30362
Qy	2851	CTATTTCATTTCTTTCCAAATGCTTTAATGATTGACGTGACAGAAAGCCAGATTCC	2910
Dp	30362	CTATTTCATTTCTTTCCAAATGCTTTAATGATTGACGTGACAGAAAGCCAGATTCC	30422
Qy	2911	TCTCGAAGCTGCAAGTATGAGCATCACTGAACTGCCGTGCGTCTCTTTTCCCTCC	2970
Dp	30422	TCTCGAAGCTGCAAGTATGAGCATCACTGAACTGCCGTGCGTCTCTTTTCCCTCC	30482
Qy	2971	TGCACGCTCATGCGGCGCTTAGCATGTTCTTTTAAATGAAGTTGTCCAGATCAGG	3030
Dp	30482	TGCACGCTCATGCGGCGCTTAGCATGTTCTTTTAAATGAAGTTGTCCAGATCAGG	30542
Qy	3031	AAAGTTCGACAAAGATATTACGAAGGTGGAACCTTACTCATCTACAGATCTCAT	3090
Dp	30542	AAAGTTCGACAAAGATATTACGAAGGTGGAACCTTACTCATCTACAGATCTCAT	30601
Qy	3091	TAGAAAGAAATATGCTAAAGCCGTGAGCATGAGCGGTGAGGAGGAAAAAGAGGACAA	3150
Dp	30602	TAGAAAGAAATATGCTAAAGCCGTGAGCATGAGCGGTGAGGAGGAAAAAGAGGACAA	30662
Qy	3151	GAAAGCTACCATTTTAAACAGTCCTTGTTATCTAGAGCAACATAATACAGCTTAAAT	3210
Dp	30662	GAAAGCTACCATTTTAAACAGTCCTTGTTATCTAGAGCAACATAATACAGCTTAAAT	30722
Qy	3211	GCACCTTAACCATGCTCGTGCGCTCCCAAAATCGTGTGCTGCTGCTGCTGGA	3270
Dp	30722	GCACCTTAACCATGCTCGTGCGCTCCCAAAATCGTGTGCTGCTGCTGCTGGA	30782
Qy	3271	CGCTTGAACGTGATGTTTGTTAGAAATCATGTTGACACCTTTTGCTCAAAAGAGCT	3330

D	30782	GCCCTGGACGTATGTTTGGTGTAAGAAATCATCTTCGACACCCTTTGTCTCAACAAGAAGCCT	30841
OY	3331	TCGTGAACACTGAGAAGAACATCTCTTTGCCATTCTGCACAGTCTCTCTACACATT	3390
Db	30842	TCTGGAACACTGAGAAGAACATCTCTTTGCCATTCTGCACAGTCTCTCTACACATT	30901
OY	3391	TTCTTCAGTCCATCTTCTGCGCTGTCTGCTCTAAGAAATTCATGGAGCTTCTCTACT	3450
Db	30902	TTCTTCAGTCCATCTTCTGCGCTGTCTGCTCTAAGAAATTCATGGAGCTTCTCTACT	30961
OY	3451	ACTAATTCAGACAGCTCTCCCTCAAAAACGTTGACTAGTCTTCAATGACCTTAACATA	3510
Db	30962	ACTAATTCAGACAGCTCTCCCTCAAAAACGTTGACTAGTCTTCAATGACCTTAACATA	31021
OY	3511	TGTACCATATACATTAATTTCAATGTTCTCCAAATAGTATTTTTTAAAGCAAATGATAC	3570
Db	31022	TGTACCATATACATTAATTTCAATGTTCTCCAAATAGTATTTTTTAAAGCAAATGATAC	31081
OY	3571	CTGTTTGGCAAAAGTAAATGATGAAGAGCTCTTAGAATTCCTCAATTTTGGCACATATCA	3630
Db	31082	CTGTTTGGCAAAAGTAAATGATGAAGAGCTCTTAGAATTCCTCAATTTTGGCACATATCA	31141
OY	3631	GTCTCTTAATATCAGAGATCCCTAAGTCAGCTGGCTAGTTACAGAGTTTTTTCAGACTT	3690
Db	31142	GTCTCTTAATATCAGAGATCCCTAAGTCAGCTGGCTAGTTACAGAGTTTTTTCAGACTT	31201
OY	3691	CCTGCTTCTCAGCTCTTATATCTTAAGACACACAGCATATATCTCTTAGAATTAACAC	3750
Db	31202	CCTGCTTCTCAGCTCTTATATCTTAAGACACACAGCATATATCTCTTAGAATTAACAC	31261
OY	3751	TAAATGGCAGTAGCGGAGATGCACACCTGACCCCGCCGCGGCGAGAGTGAGACT	3810
Db	31262	TAAATGGCAGTAGCGGAGATGCACACCTGACCCCGCCGCGGCGAGAGTGAGACT	31321
OY	3811	TTGTCTCTATTAACAAAAAGAAAAAGAAAGAAATACACCTAAGCTCA	3857
Db	31322	TTGTCTCTATTAACAAAAAGAAAAAGAAAGAAATACACCTAAGCTCA	31368
RESULT 12			
LOCUS	AC010401/C	119172 bp	DNA linear PRI 02-NOV-2001
DEFINITION	Homo sapiens chromosome 16 clone CTD-2135D7, complete sequence.		
ACCESSION	AC010401		
VERSION	AC010401.6	GI:16596527	
KEYWORDS	HTG.		
SOURCE	Homo sapiens.		
ORGANISM	Homo sapiens		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
REFERENCE	1 (bases 1 to 119172)		
AUTHORS	DOE Joint Genome Institute.		
JOURNAL	Sequencing of Human Chromosome 16		
REFERENCE	unpublished		
AUTHORS	2 (bases 1 to 119172)		
JOURNAL	Direct Submission		
REFERENCE	Submitted (15-SEP-1999) Production Sequencing Facility, DOE Joint		
AUTHORS	Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA		
JOURNAL	3 (bases 1 to 119172)		
REFERENCE	DOE Joint Genome Institute.		
AUTHORS	Submitted (02-NOV-2001) Production Sequencing Facility, DOE Joint		
JOURNAL	Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA		
COMMENT	On Nov 2, 2001 this sequence version replaced gi:19256203.		

Sequence quality assessment: This entry has been annotated with sequence quality estimates computed by the Phrap assembly program. All manually edited bases have been reduced to quality zero. Quality levels above 40 are expected to have less than 1 error in 10,000 bp.

Base by base quality values are not generally visible from the GenBank flat file format but are available as part

of this entry's ASN.1 file.

Sequence Quality Assessment:

This entry has been annotated with sequence quality estimates computed by the Phrap assembly program. All manually edited bases have been reduced to quality zero. Quality levels above 40 are expected to have less than 1 error in 10,000 bp.

Base-by-base quality values are not generally visible from the GenBank flat file format but are available as part of this entry's ASN.1 file.

FEATURES

```
Location/Qualifiers
1. .119172
   /organism="Homo sapiens"
   /db_xref="taxon:9606"
   /chromosome="16"
```

BASE COUNT	35303	a	25400	c	25120	g	33349	t
ORIGIN								

Query Match	47.4%;	Score 1829.4;	DB 9;	Length 119172;
Best Local Similarity	99.4%;	Pred. No. 0;		
Matches 1836; Conservative	0;	Mismatches 11;	Indels 0;	Gaps 0;

Oy	2011	ACGACACACGGAAGGCGCGTGGCCAGGACAGAAATGGTATACAGCAAAACATATAGTTGG	2070
Db	87961	ACTTACTTTTCTCTGTAGTGTGGCCAGGACAGAAATGGTATACAGCAAAACATATAGTTGG	87902
Oy	2071	GATATCGCGTAATTTTCGAAAGTAGACCTTCATCTCTGATTCATCTCTCGGGGCAATTGCATTT	2130
Db	87901	GATATCGCGTAATTTTCGAAAGTAGACCTTCATCTCTGATTCATCTCTCGGGGCAATTGCATTT	87844
Oy	2131	GAACCCGTGACTTTAGAGGTGGAGATTACATTCCTCACTCCAGAAATGTGCGTGGAGCCG	2190
Db	87841	GAACCCGTGACTTTAGAGGTGGAGATTACATTCCTCACTCCAGAAATGTGCGTGGAGCCG	87782
Oy	2191	AAGAAGTACGATGATTTAATCGGCTCTTAAATTAACGGCCGCGCTCTACACCAAGTGCATTC	2250
Db	87781	AAGAAGTACGATGATTTAATCGGCTCTTAAATTAACGGCCGCGCTCTACACCAAGTGCATTC	87722
Oy	2251	TTCATGTCCCGCTACTACACAAGCGTCCGCTGCTTCGATTAAGTTTGACCCCTACCAAGCCT	2310
Db	87721	TTCATGTCCCGCTACTACACAAGCGTCCGCTGCTTCGATTAAGTTTGACCCCTACCAAGCCT	87662
Oy	2311	TTCTCTCACTTCCCGAGGTGCCCTTGTTTCAGAGATCTTCACAGCAATGACATTAAGTTCC	2370
Db	87661	TTCTCTCACTTCCCGAGGTGCCCTTGTTTCAGAGATCTTCACAGCAATGACATTAAGTTCC	87602
Oy	2371	AAACTACATCTTCTTACACTTACTTCCCGACAGTACGGGATCTCTGTGTCGCCCTCTCCT	2430
Db	87601	AAACTACATCTTCTTACACTTACTTCCCGACAGTACGGGATCTCTGTGTCGCCCTCTCCT	87542
Oy	2431	CATGCAACGGCGGAGTGTGTTGAGGAGTGGAAACAAGCAAGCCACAGCCTGATGGGGGG	2490
Db	87541	CATGCAACGGCGGAGTGTGTTGAGGAGTGGAAACAAGCAAGCCACAGCCTGATGGGGGG	87482
Oy	2491	ACAGCACTGGCCATTACAGACAGATCTGAAACCCCTGCCAGTCTAGAGAGTAAATCTCT	2550
Db	87481	ACAGCACTGGCCATTACAGACAGATCTGAAACCCCTGCCAGTCTAGAGAGTAAATCTCT	87422
Oy	2551	GGTCCCAAGACTCTTGTGTTAAAAATGCGAGGGGTGAATGGCAAAAATCGCGCTCTCTTG	2610
Db	87421	GGTCCCAAGACTCTTGTGTTAAAAATGCGAGGGGTGAATGGCAAAAATCGCGCTCTCTTG	87362
Oy	2611	ATGCAACCAAGTTTAAGAAACATCGCAGAAATTACAGACCGCTGTACCAAGACAGCTACAGAGT	2670
Db	87361	ATGCAACCAAGTTTAAGAAACATCGCAGAAATTACAGACCGCTGTACCAAGACAGCTACAGAGT	87302
Oy	2671	ATTCTGGGGATGCTGCAAAATGCCAAACAGCTTTATGATTTTCATTCACACCTTTTGGG	2730
Db	87301	ATTCTGGGGATGCTGCAAAATGCCAAACAGCTTTATGATTTTCATTCACACCTTTTGGG	87242

QY	2731	GAAGCGATATCAAAAGGAAAAAGGAAAAAGTCAACAGATGAGCGCTTTTCTTATCCAT	2790
Db	87241	GAACTCGATATCAAAAGGAAAAAGGAAAAAGTACAGTATGATGCTTTTCTTATCCAT	87182
QY	2791	GCGTGTACTTTTACAGGGGCTCTTCCAGACATCATAGTGCATTTATATATATGTTG	2850
Db	87181	GCGTGTACTTTTACAGGGGCTCTTCCAGACATCATAGTGCATTTATATATATGTTG	871222
QY	2851	CTATTTCATTTCTTTTCCAAATGCTCTTAATGATTTGACGGTGGACCAAGAAGCAGATTCC	2910
Db	87121	CTATTTCATTTCTTTTCCAAATGCTCTTAATGATTTGACGGTGGACCAAGAAGCAGATTCC	87062
QY	2911	TCTCTGACCTCGCAGTTAGGATCATCTGAACTGCGGTGCGGCTTTTCTCTCC	2970
Db	87061	TCTCTGAACTCTGCAGTTAGGATCATCTGAACTGCGGTGCGGCTTTTCTCTCC	87002
QY	2971	TGCACGCTATATGCGGGGCTTAGCATGTTCTTTTAAATGAGTGTGTCCAGATCAGGT	3030
Db	87001	TGCACCGCTATATGCGGGGCTTAGCATGTTCTTTTAAATGAGTGTGTCCAGATCAGGT	86942
QY	3031	AAAGTTCTTACAAAGTATACAGAGTAAAGAACTTTACCTGATCTTACAGATCTCAT	3090
Db	86941	AAAGTTCTTACAAAGTATACAGAGTAAAGAACTTTACCTGATCTTACAGATCTCAT	86882
QY	3091	TAGAAAGCAATATGTCTAGACCTGGCATGTGACGGTGCAGGAGGAAAGACAGGCACAA	3150
Db	86881	TAGAAAGCAATATGTCTAGACCTGGCATGTGACGGTGCAGGAGGAAAGACAGGCACAA	86822
QY	3151	GAAACCTACCAATTTTAAACAGCCCTGTATATCTAGCAACATATATACAGTCTTAAT	3210
Db	86821	GAAACCTACCAATTTTAAACAGCTCTTGTATCTAGTGCACATATATACAGTCTTAAT	86762
QY	3211	GCACCTATATACCATGTCTGTGGCTCTCCAAATCTGCTTGTGTGTGTCTCTGGA	3270
Db	86761	GCACCTATATACCATGTCTGTGGCTCTCCAAATCTGCTTGTGTGTGTCTCTGGA	86702
QY	3271	CGCTTGAACATGATGTTTGTGTAGAGAAATCATGTCTGACCTTTGTCTTACAAAGAGCT	3330
Db	86701	CGCTTGAACATGATGTTTGTGTAGAGAAATCATGTCTGACCTTTGTCTTACAAAGAGCT	86642
QY	3331	TCTGGAACACAGAGAGAAACATCTCTTGGCATCTCGACACATCTCTCACACAT	3390
Db	86641	TCTGGAACACAGTACAGAGAAACATCTCTTGGCATCTCGACACATCTCTCACACAT	86582
QY	3391	TTCTTCACGCTCCATCTCTGCTGTCTGCTCTAAGAAATTCATGAGCCTTCTACT	3450
Db	86581	TTCTTCACGCTCCATCTCTGCTGTCTGCTCTAAGAAATTCATGAGCCTTCTACT	86522
QY	3451	ACTAATTTAAGACAGCTCTCTCTCAAAACTGGTGTGCTAGTCTCTAATGACCCTAACATA	3510
Db	86521	ACTAATTTAAGACAGCTCTCTCTCAAAACTGGTGTGCTCTCTAATGACCCTAACATA	86462
QY	3511	TGTACCATATCTATAAATTTCTTCAATTTGTCCTAATTTTAAAGCAAAATGAAATAC	3570
Db	86461	TGTACCATATCTATAAATTTCTTCAATTTGTCCTAATTTTAAAGCAAAATGAAATAC	86402
QY	3571	CTGTTTGCAGAAATTAATGATGAGAGAGCTTTAGAAATTCATATTTTTCACATATTTCA	3630
Db	86401	CTGTTTGCAGAAATTAATGATGAGAGAGCTTTAGAAATTCATATTTTTCACATATTTCA	86342
QY	3631	GTCCTCTAATATCAGAGATCCCTTAAGTCCAGCTGGCTAGTACAGATTTTTCAGACTT	3690
Db	86341	GTCCTCTAATATCAGAGATCCCTTAAGTCCAGCTGGCTAGTACAGATTTTTCAGACTT	86282
QY	3691	CCTGCTTCTCAGCTCTTATATCTTAATCCTAAGACACACACATCATATCTCTTAGAATATCAAC	3750
Db	86281	CCTGCTTCTCAGCTCTTATATCTTAATCCTAAGACACACACATCATATCTCTTAGAATATCAAC	86222
QY	3751	TAAATGGCAGTAGGCGGAGATCGACCACTGCAACCCCTGCTGGCGCAGACAGATGAGACT	3810
Db	86221	TAAATGGCAGTAGGCGGAGATCGACCACTGCAACCCCTGCTGGCGCAGACAGATGAGACT	86162
QY	3811	TTGTCTCTAATTCAAAAAGAAAGAAATACACCTTAAGCTCA 3857	

|||||

Db 86161 TTGTCCTATACAAAAGAAAAGAAATACAACTAGCTCA 86115

RESULT 13
AC009173 149849 bp DNA linear PRI 26-APR-2002
DEFINITION Homo sapiens chromosome 16 clone RP11-99H5, complete sequence.
AC009173
VERSION AC009173.8 GI:20330796
KEYWORDS HRC.
SOURCE
ORGANISM Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
1 (bases 1 to 149849)
DOE Joint Genome Institute.
Sequencing of Human Chromosome 16
Unpublished
2 (bases 1 to 149849)
DOE Joint Genome Institute.
Direct Submission
Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
3 (bases 1 to 149849)
DOE Joint Genome Institute.
Direct Submission
Submitted (06-MAR-2002) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
4 (bases 1 to 149849)
DOE Joint Genome Institute.
Direct Submission
Submitted (26-APR-2002) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
On Apr 26 2002 this sequence version replaced gi:119172832.

COMMENT
Sequence Quality Assessment:
This entry has been annotated with sequence quality
estimates computed by the Phrap assembly program.
All manually edited bases have been reduced to quality zero.
Quality levels above 40 are expected to have less than
1 error in 10,000 bp.
Base-by-base quality values are not generally visible from the
GenBank flat file format but are available as part
of this entry's ASN.1 file.

Sequence Quality Assessment:
This entry has been annotated with sequence quality
estimates computed by the Phrap assembly program.
All manually edited bases have been reduced to quality zero.
Quality levels above 40 are expected to have less than
1 error in 10,000 bp.
Base-by-base quality values are not generally visible from the
GenBank flat file format but are available as part
of this entry's ASN.1 file.

FEATURES
source Location/Qualifiers
1. 149849
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="16"
/clone="RP11-99H5"
BASE COUNT 43996 a 31384 c 30828 g 43641 t
ORIGIN

Query Match 47.3%; Score 1826.2; DB 9; Length 149849;
Best Local Similarity 99.3%; Pred. No. 0;
Matches 1834; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 2011 ACTGACACCTCGGAAGCCGGTGGCCAGAGCAAGATGTGTACACAGCAAGCATAGTTGTG 2070
|||||
Db 115652 ACTTATCTTCTCTGTGTGTGGTGGCCAGGAAGCAATGTGTACACAGCAAGCATAGTTGTG 115711

QY 2071 GATATGCGTAATTTGGAAGTGAAGCTTCATCTGTATCCATCTGTCGGGGCATTTGACATT 2130

|||||

Db 115712 GATATGCGTAATTTGGAAGTGAAGCTTCATCTGTATCCATCTGTCGGGGCATTTGACATT 115771

QY 2131 GAACCCGTGACTTTAGAGTTGGAGATTACATCTGACTCCAGAAATGTGCGTGAGCCG 2190
|||||
Db 115772 GAACCCGTGACTTTAGAGTTGGAGATTACATCTGACTCCAGAAATGTGCGTGAGCCG 115831

QY 2191 AAGAGTATGATGATTTAATTCGGCTTTAATTAATACGGCGGCTCTACAGCAAGTGCATC 2250
|||||
Db 115832 AAGAGTATGATGATTTAATTCGGCTTTAATTAATACGGCGGCTCTACAGCAAGTGCATC 115891

QY 2251 TCCATGTCCCGCTACTACAAAGGTCGCCGTCTTGTGATTTGATGACCAAGCCCT 2310
|||||
Db 115892 TCCATGTCCCGCTACTACAAAGGTCGCCGTCTTGTGATTTGATGACCAAGCCCT 115951

QY 2311 TTCTCTCTACTTCCCGAGAGTGCCTTTGTTACAGAGATCTCCAGCAATGACATTAATTC 2370
|||||
Db 115952 TTCTCTCTACTTCCCGAGAGTGCCTTTGTTACAGAGATCTCCAGCAATGACATTAATTC 116011

QY 2371 AAACCTACTCTTCTTACACTTCCAGCTTCCAGACTTCTGCTGCTGCTGCTGCTGCT 2430
|||||
Db 116012 AAACCTACTCTTCTTACACTTCCAGACTTCCAGACTTCTGCTGCTGCTGCTGCTGCT 116071

QY 2431 CATGCAACGCGGAGTTGTTGAGAGCTGAACAAAGCAAGCCACAGCCATGATGCGGCG 2490
|||||
Db 116072 CATGCAACGCGGAGTTGTTGAGAGCTGAACAAAGCAAGCCACAGCCATGATGCGGCG 116131

QY 2491 ACAGCACTGGCCATTTACAGACATTTGAAACCCCTCCGAGTCAAGAAATTAATTCCT 2550
|||||
Db 116132 ACAGCACTGGCCATTTACAGACATTTGAAACCCCTCCGAGTCAAGAAATTAATTCCT 116191

QY 2551 GGTCCCAAGACTCTTGTGTTAAATGCCAGGGGTAAATGCCAAATGCGGCTCTTG 2610
|||||
Db 116192 GGTCCCAAGACTCTTGTGTTAAATGCCAGGGGTAAATGCCAAATGCGGCTCTTG 116251

QY 2611 ATGCACACGTTAAGAAATCGCAGATTAGCAGCCCTGCACAGACGAGCTCAGAGCT 2670
|||||
Db 116252 ATGCACACGTTAAGAAATCGCAGATTAGCAGCCCTGCACAGACGAGCTCAGAGCT 116311

QY 2671 ATTCTGGGGAATGCTGCAATATGCAAAAGCCTTAATGATTTCAATACACCTTTTGA 2730
|||||
Db 116312 ATTCTGGGGAATGCTGCAATATGCAAAAGCCTTAATGATTTCAATACACCTTTTGA 116371

QY 2731 GAAGTGTATCAAAAGAAAGGAAAGAAAGTGAACAGTATGATGCTTTTCTATCCAT 2790
|||||
Db 116372 GAAGTGTATCAAAAGAAAGGAAAGAAAGTGAACAGTATGATGCTTTTCTATCCAT 116431

QY 2791 GCCTGTACTTTTCAGCGGCTCCTTGCAGACATCATAGTCAATTAATTAATTTGCTTG 2850
|||||
Db 116432 GCCTGTACTTTTCAGCGGCTCCTTGCAGACATCATAGTCAATTAATTAATTTGCTTG 116491

QY 2851 CTAATTCATTTCTTTTCAATGCTCTTAATGATTTGTACGTGACGCAAGCCAGATTC 2910
|||||
Db 116492 CTAATTCATTTCTTTTCAATGCTCTTAATGATTTGTACGTGACGCAAGCCAGATTC 116551

QY 2911 TCTCTGAACCTGCAATGATGAGATCACTGCACTGCGCTGCTGCTGCTGCTGCTGCT 2970
|||||
Db 116552 TCTCTGAACCTGCAATGATGAGATCACTGCACTGCGCTGCTGCTGCTGCTGCTGCTGCT 116611

QY 2971 TGCACCGCTATGCGCGGCTTACGATGTTCTTTTAAATGAGGTTGTACAGATCAGGT 3030
|||||
Db 116612 TGCACCGCTATGCGCGGCTTACGATGTTCTTTTAAATGAGGTTGTACAGATCAGGT 116671

QY 3031 AAAGTCTCAAGAGTATTACAGAAAGTAGAAATTTTACCTGATCTTAACAGATCTCAAT 3090
|||||
Db 116672 AAAGTCTCAAGAGTATTACAGAAAGTAGAAATTTTACCTGATCTTAACAGATCTCAAT 116731

QY 3091 TGAAGAGGAATATGCTTAACCTTGCAATGACGCTGACGAGGAGGAGGAGGAGGAGGAGCA 3150
|||||
Db 116732 TGAAGAGGAATATGCTTAACCTTGCAATGACGCTGACGAGGAGGAGGAGGAGGAGGAGCA 116791

QY 3151 GAAAGCTACATTTTAAAGTCTGTTATCTAGTGCAACATTAATTAACAGTCTTAAT 3210
|||||

Db 116792	GAAAGCTACCATTTTAAAGCTCTGTTATCTAGTCACATTAATTAACAGTCTAATT	116851	variation	122	/frequency="0.01"
Qy 3211	GCATTATACCATGCTCTGCTGCTCCCAATCTGGCTCTTGGCTGTGCTGCTGGA	3270	repeat_region	129	/replace="c"
Db 116852	GCATTATACCATGCTCTGCTGCTCCCAATCTGGCTCTTGGCTGTGCTGCTGGA	116911	variation	248	/rpt_family="L1"
Qy 3271	CGCTTGAACGATCTTGTGTAGGAATCATGTTCTGACCTTTGTCTACAAAGAGCCT	3330	variation	248	/rpt_type-dispersed
Db 116912	CGCTTGAACGATCTTGTGTAGGAATCATGTTCTGACCTTTGTCTACAAAGAGCCT	116971	variation	248	/frequency="0.03"
Qy 3331	TCGTGAACACTGAGAGAACAATCTCTTGGCAATCTCTGACCACTCTCTCTACACATT	3390	variation	424	/replace="g"
Db 116972	TCGTGAACACTGAGAGAACAATCTCTTGGCAATCTCTGACCACTCTCTCTACACATT	117031	repeat_region	450	/frequency="0.29"
Qy 3391	TTCTTCAGCTCCATCTCTTCTGCTGCTGCTCTAGGAATTTCTAGAGCTTCTACT	3450	variation	460	/replace="t"
Db 117032	TTCTTCAGCTCCATCTCTTCTGCTGCTGCTCTAGGAATTTCTAGAGCTTCTACT	117091	variation	460	/rpt_family="Alu"
Qy 3451	ACTAATTCAGACAGCTCTCTCCCAAAACGTGGTACTAGTCTTCTAATGACCTTAACATA	3510	variation	892	/rpt_type-dispersed
Db 117092	ACTAATTCAGACAGCTCTCTCCCAAAACGTGGTACTAGTCTTCTAATGACCTTAACATA	117151	variation	975	/frequency="0.01"
Qy 3511	TGTAGCATATATCTAATTAATTCATGTTCCAAATTAATTAATTAAGCAAAATGAATTAC	3570	variation	1038	/replace="g"
Db 117152	TGTAGCATATATCTAATTAATTCATGTTCCAAATTAATTAATTAAGCAAAATGAATTAC	117211	variation	1038	/replace="c"
Qy 3571	CTGTTTCAAAAGTTAATGATGAAGGAGCTTTGATTCATCTTCTGACATATTCATA	3630	gene	<1068	/frequency="0.22"
Db 117212	CTGTTTCAAAAGTTAATGATGAAGGAGCTTTGATTCATCTTCTGACATATTCATA	117271	mRNA	<1068	/replace="a"
Qy 3631	GTCCTCTAATATATCAGAGATCCCTAAGTCAGCTGGCTAGTATACAGAGTTTTCAGACTT	3690	variation	11619	/gene="ERCC4"
Db 117272	GTCCTCTAATATATCAGAGATCCCTAAGTCAGCTGGCTAGTATACAGAGTTTTCAGACTT	117331	CDS	11619	/product="excision repair cross-complementing rodent repair deficiency, complementation group 4"
Qy 3691	CCTGCTTCTCAGCTCTTATATCTAGACACCGACATCATATCTCTAGAAATACAAAC	3750	variation	11619	/frequency="0.01"
Db 117332	CCTGCTTCTCAGCTCTTATATCTAGACACCGACATCATATCTCTAGAAATACAAAC	117391	variation	18674	/replace="g"
Qy 3751	TAAATGACAGTGAAGCCAGATCGACACCTGCTGCTGGGCGACAGAGTGAAGACT	3810	variation	18674	/replace="c"
Db 117392	TAAATGACAGTGAAGCCAGATCGACACCTGCTGCTGGGCGACAGAGTGAAGACT	117451	variation	18674	/frequency="0.01"
Qy 3811	TTGCTCTATTTACAAAAGAAAAGAAATACAACTTAAGCTCA	3857	variation	18674	/replace="a"
Db 117452	TTGCTCTATTTACAAAAGAAAAGAAATACAACTTAAGCTCA	117498	variation	18674	/replace="c"
RESULT 14					
AF491814					
LOCUS	29755 bp	DNA	linear	PRI 20-MAR-2002	
DEFINITION	Homo sapiens excision repair cross-complementing rodent repair deficiency, complementation group 4 (ERCC4) gene, complete cds.				
ACCESSION	AF491814				
VERSION	AF491814.1				
KEYWORDS	GI:19550956				
SOURCE	Homo sapiens.				
ORGANISM	Homo sapiens				
REFERENCE	1 (bases 1 to 29755)				
AUTHORS	Rieder,M.J., Braun,A.C., Montoya,M.A., Chung,M.-W., Nguyen,C.P., Nguyen,D.A., Livingston,R.J., Poel,C.L., Robertson,P.D., Schackwitz,W.S., Sherwood,J.K., Wiltrik,L.A. and Nickerson,D.A.				
TITLE	Submitted (12-MAR-2002) Genome Sciences, University of Washington, 1705 NE Pacific, Seattle, WA 98195, USA				
JOURNAL	Direct Submission				
COMMENT	To cite this work please use: NIHES-SNPs, Environmental Genome Project, NIHES ES15478, Department of Genome Sciences, Seattle, WA (URL: http://egp.gs.washington.edu).				
FEATURES	Location/Qualifiers				
source	1..29755				
	/organism="Homo sapiens"				
	/db_xref="taxon:9606"				

variation	1406	/gene="ERCC4"	/frequency="0.99"	/replace="n"
variation	1421	/gene="ERCC4"	/frequency="0.01"	/replace="t"
variation	1556	/gene="ERCC4"	/frequency="0.01"	/replace="c"
variation	1757	/gene="ERCC4"	/frequency="0.01"	/replace="g"
variation	2016	/gene="ERCC4"	/frequency="0.01"	/replace="a"
variation	2127	/gene="ERCC4"	/frequency="0.21"	/replace="c"
variation	2152	/gene="ERCC4"	/frequency="0.01"	/replace="g"
variation	2298	/gene="ERCC4"	/frequency="0.04"	/replace="a"
variation	2336	/gene="ERCC4"	/frequency="0.02"	/replace="t"
variation	2343	/gene="ERCC4"	/frequency="0.01"	/replace="c"
variation	2356	/gene="ERCC4"	/frequency="0.01"	/replace="g"
variation	2358	/gene="ERCC4"	/frequency="0.01"	/replace="a"
variation	2359	/gene="ERCC4"	/frequency="0.01"	/replace="t"
variation	2633	/gene="ERCC4"	/frequency="0.03"	/replace="g"
variation	2653	/gene="ERCC4"	/frequency="0.01"	/replace="t"
variation	2978	/gene="ERCC4"	/frequency="0.03"	/replace="g"
variation	3241	/gene="ERCC4"	/frequency="0.01"	/replace="t"
variation	3258	/gene="ERCC4"	/frequency="0.01"	/replace="t"
variation	3528	/gene="ERCC4"	/frequency="0.01"	/replace="a"
variation	3538	/gene="ERCC4"	/frequency="0.01"	/replace="a"

```
variation
/frequency="0.01"
/replace="a"
3559
/genome="ERCc4"
/frequency="0.01"
/replace="a"
3628..3772
/rpl_family="L2"
/rpl_type=dispersed
4047
variation
/genome="ERCc4"
/frequency="0.01"
/replace="t"
4095..4261
/rpl_family="L2"
/rpl_type=dispersed
4332
variation
/genome="ERCc4"
/frequency="0.73"
/replace="n"
4338
variation
/genome="ERCc4"
/frequency="0.01"
/replace="g"
4429
/genome="ERCc4"
/frequency="0.26"
/replace="t"
4436
/genome="ERCc4"
```

Query Match	32.0%;	Score 1233.4;	DB 9;	Length 29755;	.
Best Local Similarity	99.1%;	Pred. No. 2.5e-282;		:	
Matches 1240;	Conservative 0;	Mismatches 11;	Indels 0;	Gaps 0;	

QY	2011	ACGACACCTCGGAACACCCCGTGGCCAGGAACAGAAATGTRACACAGCAACAAAGCATAGTTGTG	2070
Db	28505	ACTTACTTTTCTCTGTAGGTGGCCAGGAACAAAGTGTACACAGCAAAAGCATAGTTGTG	2856
QY	2071	GATATCGGTGAATTTGGAAGTGAAGTGCATCTGTATCATCTGTGCAGGTCATTTGCACAT	2130
Db	28565	GATATCGGTGAATTTGGAAGTGAAGTGCATCTGTATCATCTGTGCAGGTCATTTGCACAT	28622
QY	2131	GAACCCGTACTTTAGAGGTGGAGATTACATCTCCTACCTCAGAAATGTGCGTGAAGCG	2190
Db	28625	GAACCCGTACTTTAGAGGTGGAGATTACATCTCCTACCTCAGAAATGTGCGTGAAGCG	28688
QY	2191	AAGAGATGAGATTTATTCGGCTTTTAAATAAGGGCGCTCTACAGCCAGTGCATC	2250
Db	28685	AAGAGATGAGATTTATTCGGCTTTTAAATAAGGGCGCTCTACAGCCAGTGCATC	28744
QY	2251	TCCATGTCCCGGTACACACAAGGTCGCCGTCTGTATTGAGTTGACCTTACCAAGCCT	2310
Db	28745	TCCATGTCCCGGTACACACAAGGTCGCCGTCTGTATTGAGTTGACCTTACCAAGCCT	28804
QY	2311	TTCCTCTACTTTCCGAGGTCCTTGTTCAGAGATCTCCAGCAATGACATTAGTTCC	2370
Db	28805	TTCCTCTACTTTCCGAGGTCCTTGTTCAGAGATCTCCAGCAATGACATTAGTTCC	28864
QY	2371	AAACACTACTTCTTAACTTCATCTCCCGAGACTAGGATTTCTCTGAGGCCCTTCCT	2430
Db	28865	AAACACTACTTCTTAACTTCATCTCCCGAGACTAGGATTTCTCTGAGGCCCTTCCT	28922
QY	2431	CATGCAAGGGCGAGTTGTTTGAAGAGCTGAACAACAAAGCAACGACGCTGATGCGGG	2490
Db	28925	CATGCAAGGGCGAGTTGTTTGAAGAGCTGAACAACAAAGCAACGACGCTGATGCGGG	28988
QY	2491	ACAGACAGGGCATTTACACACATTCGTGAACCTTCGCGAGCTCAGAGAGATATTCCT	2550
Db	28985	ACAGACAGGGCATTTACACACATTCGTGAACCTTCGCGAGCTCAGAGAGATATTCCT	29044
QY	2551	GGTCCCAAGACTTTTGTAAATAATGCCAGGGGTGAATGCCAAATACGCGCTCTTG	2610

Db	29045	GCCTCCCAAGACCTCTTGTTAAAAATGCGACAGGGGTGAATGCCAAAAAACTGCCGCTCTTG	291104
QY	2611	ATGCACACAGTTTAAGAACATCGCAGAAATTAGCAGCCCTGTACAAAGACGACCTACAGAGT	2670
Db	29105	ATGCACACAGTTTAAGAACATCGCAGAAATTAGCAGCCCTGTACAAAGACGACCTACAGAGT	29164
QY	2671	ATTCTGGGGAATTCGTCGCAAAATGCCAAACAGGGTTATGATTTTCATTCACACCCCTTTTCA	2730
Db	29165	ATTCTGGGGAATTCGTCGCAAAATGCCAAACAGGGTTATGATTTTCATTCACACCCCTTTTCA	29224
QY	2731	GAACCTGTATCAAAAAGGAAAAAGGAAAAATGAAACAGTATGAGCTGTTTCTTATCCAT	2790
Db	29225	GAACCTGTATCAAAAAGGAAAAAGGAAAAATGAAACAGTATGAGCTGTTTCTTATCCAT	29284
QY	2791	GCCGTACTTTTCAGCGGCTCCTTGCCAGAAATCATAGTATATTATTAATTATGGTTTG	2850
Db	29285	GCCGTACTTTTCAGCGGCTCCTTGCCAGAAATCATAGTATATTATTAATTATGGTTTG	29344
QY	2851	CTATTTCATCTTTTCCAAATGCTCTTAATGATTTGACGGTGGACAGCAAGACAGATTC	2910
Db	29345	CTATTTCATCTTTTCCAAATGCTCTTAATGATTTGACGGTGGACAGCAAGACAGATTC	29404
QY	2911	TCCTGAACTCTGCAAGTTAGGCAATCACTTGAACCTTGCCGTGCTCTTTTCTCC	2970
Db	29405	TCCTGAACTCTGCAAGTTAGGCAATCACTTGAACCTTGCCGTGCTCTTTTCTCC	29464
QY	2971	TGCACCCGCTATGCGCGGCTTAGCATGTTCTTTTAAATAGGTTTGCAGAGATCAGT	3030
Db	29465	TGCACCCGCTATGCGCGGCTTAGCATGTTCTTTTAAATAGGTTTGCAGAGATCAGT	29524
QY	3031	AAAGTTCCTACAGATGATTAAGAAAGTAAAGAACTTACCTGATCTTAACAGATCTCAT	3090
Db	29525	AAAGTTCCTACAGATGATTAAGAAAGTAAAGAACTTACCTGATCTTAACAGATCTCAT	29584
QY	3091	TAGAAAGATATGCTAAGCTGCGCATGAGCGTGCAGAGGAGGAAAAAGCAGGCACAA	3150
Db	29585	TAGAAAGATATGCTAAGCTGCGCATGAGCGTGCAGAGGAGGAAAAAGCAGGCACAA	29644
QY	3151	GAAAGCTACCAATTTTAAACAGTCCCTGTTATCTAGTGAACAATAAATTAACAGTCTTAAT	3210
Db	29645	GAAAGCTACCAATTTTAAACAGTCCCTGTTATCTAGTGAACAATAAATTAACAGTCTTAAT	29704
QY	3211	GCACCTATATACCCATGCTCCTGTGGCTCTCCAAATCGCTTGTGCTGTGTG	3261
Db	29705	GCACCTATATACCCATGCTCCTGTGGCTCTCCAAATCGCTTGTGCTGTGTG	29755
RESULT 15			
LOCUS	ARI813108	1101 bp	DNA
DEFINITION	Sequence 12 from patent US 6340566.		
ACCESSION	ARI813108		
VERSION	ARI813108.1	GI:20226701	
KEYWORDS	Unknown.		
SOURCE	Unknown.		
ORGANISM	Unknown.		
REFERENCE	1 (bases 1 to 1101)		
AUTHORS	McCutchen-Maloney,S.L.		
TITLE	Detection and quantitation of single nucleotide polymorphisms, DNA		
JOURNAL	sequence variations, DNA mutations, DNA damage and DNA mismatches		
FEATURES	Patent: US 6340566-A 12-22-JAN-2002;		
source	Location/Qualifiers		
	1..1101		
	/organism="unknown"		
BASE COUNT	324 a 230 c 246 g 301 t		
ORIGIN			
Query Match	28.5%; Score 1101; DB 6; Length 1101;		
Best Local Similarity	100.0%; Pred. No. 4,9e-251;		
Matches 1101; Conservative	0; Mismatches 0; Indels 0; Gaps 0;		
QY	46	ATGCGCGCGCTGCTGAGATACGAGGACAGCTGTGCTGGAACCTGCTGACACTGACGGG	105

D	b	1	ATGCGCGCGCTCGTGGAGTAGACGGACAGCTGGTGGTGGAACTGCTCGACACTGACGGG	60
Q	y	106	CTGTAGTAGTGTGCGCCCGCGGGGCTCGGCGGGGACCGGGCTCTCTACCACTTCTCCAGTG	165
D	b	61	CTAGTAGTGTGCCCCCGCGGGGCTCGGCGGGGACCGGGCTCTCTACCACTTCTCCAGTG	120
Q	y	166	CATGCGACCCACCTGCGCTGGGTGGTGGTGTCAACACGACGACCGCGCGGACGAGAGAT	225
D	b	121	CAGTGCACCCACCTGCGCTGGGTGGTGGTGTCAACACGACGACCGCGCGGACGAGAGAT	180
Q	y	226	TTTATCAATCAGCTGAAGATAGAAAGGATTTGAACACTTCCTCGCGGTCTAACAAATGAA	285
D	b	181	TTTATCAATCAGCTGAAGATAGAAAGGATTTGAACACTTCCTCGCGGTCTAACAAATGAA	240
Q	y	286	ATCACAAAGCAACAGTGGCTATGAAAGTTTACACAAAGGTGGTGTATATTGGCAGCAAGT	345
D	b	241	ATCACAAAGCAACAGTGGCTATGAAAGTTTACACAAAGGTGGTGTATATTGGCAGCAAGT	300
Q	y	346	AGGATACCTTGGGTGACCTTGACCTGTAGAAATACCTTCAGATTATTAATCTGGCATC	405
D	b	301	AGGATACCTTGGGTGACCTTGACCTTGACCTGTAGAAATACCTTCAGATTATTAATCTGGCATC	360
Q	y	406	TTGGGTATAGAGCCACAGAAATATCGAGTCTGTTCAGAAAGCATTCATCTTGGGCTTC	465
D	b	361	TTGGGTATAGAGCCACAGAAATATCGAGTCTGTTCAGAAAGCATTCATCTTGGGCTTC	420
Q	y	466	TTTTCGCCAGAAAAACAACGTGGTTTATTTAAAGCTTTCACAGCAATAGCTGGTCCCTT	525
D	b	421	TTTTCGCCAGAAAAACAACGTGGTTTATTTAAAGCTTTCACAGCAATAGCTGGTCCCTT	480
Q	y	526	GATACCTGTTTTTGTCACTGTGAAAGAGTATGAGAAATCTTTTGTG:AGGAAACTGTAT	585
D	b	481	GATACCTGTTTTTGTCACTGTGAAAGAGTATGAGAAATCTTTTGTGAGAAACTGTAT	540
Q	y	586	CTGTGGCCAAAGTTCATGTAGCACTAATCTATTTTGAACAGCACCAACCTGAAGTT	645
D	b	541	CTGTGGCCAAAGTTCATGTAGCACTAATCTATTTTGAACAGCACCAACCTGAAGTT	600
Q	y	646	GTAACAAATCATGTTCTATAGACACTACACTGCTGTGATACAGCTGCTTACTGTGAC	705
D	b	601	GTAACAAATCATGTTCTATAGACACTACACTGCTGTGATACAGCTGCTTACTGTGAC	660
Q	y	706	ATTTTAATGCATGTCTTAAAGGAAGTAAATGCGATTAACCATCGCTTGAAGTGAAGAT	765
D	b	661	ATTTTAATGCATGTCTTAAAGGAAGTAAATGCGATTAACCATCGCTTGAAGTGAAGAT	720
Q	y	766	TTATCTTTAGAAAAATGCTATTTGAAAAACCTTTTGACAAAGCAATCCGCCATTAATGAT	825
D	b	721	TTATCTTTAGAAAAATGCTATTTGAAAAACCTTTTGACAAAGCAATCCGCCATTAATGAT	780
Q	y	826	CCTTGTGGGACACGAGCTTGGAGCCAAAGACTAAATCCTTAGTTACAGATTTGAAGTATTTA	885
D	b	781	CCTTGTGGGACACGAGCTTGGAGCCAAAGACTAAATCCTTAGTTACAGATTTGAAGTATTTA	840
Q	y	886	CGAACTTTGCGAGCTATCTCTCTCAGTATGATTTGTGCACATTTCTTAACTTCTGAAA	945
D	b	841	CGAACTTTGCGAGCTATCTCTCTCAGTATGATTTGTGTGCACATTTCTTAACTTCTGAAA	900
Q	y	946	TCTCTGAGAGCAACGSAAAAAAGCTTTTGTGCAATTCAGGTTGGCTGTTCTTGACTTC	1005
D	b	901	TCTCTGAGAGCAACGSAAAAAAGCTTTTGTGCAATTCAGGTTGGCTGTTCTTGACTTC	960
Q	y	1006	AGCACCTCGATGTTTATTTAAATGCTCGAGCAGAGGGTTTATCATCTTTCAGATCCCAAAATG	1065
D	b	961	AGCACCTCGATGTTTATTTAAATGCTCGAGCAGAGGGTTTATCATCTTTCAGATCCCAAAATG	1020
Q	y	1066	AGTAAAAAGAAAAAATATCTGAAAAAATGGAATTTAAAGAACGGGAAGCAACAAAAAAG	1125
D	b	1021	AGTAAAAAGAAAAAATATCTGAAAAAATGGAATTTAAAGAACGGGAAGCAACAAAAAAG	1080
Q	y	1126	GAACGTGGCTCAGAAAGCAAC	1146

Wed Nov 27 08:49:07 2002

us-09-919-497-14.rge

Page 25

Db 1081 GACTGTCCTAGAAAGCAC 1101

Search completed: November 26, 2002, 22:46:38
Job time : 8482.76 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen, Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 26, 2002, 19:42:45 : Search time 540.781 Seconds
(without alignments)
16061.850 Million cell updates/sec

Title: US-09-919-497-14

Perfect score: 3857
Sequence: 1 ggaagagcttcacatggagtc.....agaatacaacctaagctca 3857

Scoring table: IDENTITY_NNC
Gapop 10.0, Gapext 1.0

Searched: 2185239 segs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_101002.*
1: /SID52/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SID52/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SID52/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SID52/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SID52/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SID52/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SID52/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SID52/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SID52/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SID52/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SID52/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SID52/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SID52/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SID52/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SID52/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SID52/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SID52/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SID52/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SID52/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SID52/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SID52/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SID52/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SID52/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SID52/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3857	100.0	3857	24	ABK35494
2	2875	74.5	2890	18	AA197897
3	2857.4	74.1	2900	21	AA16235
4	2718	70.5	2718	22	AA563230
5	1829.4	47.4	35641	24	ABN95727
6	1829.4	47.4	35641	24	AB164428
7	1101	28.5	1101	22	AA563229
8	256.2	6.6	601	24	AB167494
9	253.2	6.6	2995	23	AB106559

10	159	4.1	5313	23	AB106558	Drosophila melanog
C 11	154.6	4.0	3857	24	ABK35494	Human endometrial
C 12	154.6	4.0	35641	24	ABN95727	Gene #2225 used to
C 13	154.6	4.0	35641	24	AB164428	Stomach cancer rel
C 14	135.6	3.5	574	24	ABQ21888	Oligonucleotide fo
C 15	135.6	3.5	574	24	ABQ21889	Oligonucleotide fo
C 16	135.6	3.5	574	24	ABQ32918	Oligonucleotide fo
C 17	135.6	3.5	574	24	ABQ32919	Oligonucleotide fo
C 18	135.6	3.5	6866	24	AB149319	Human polynucleoti
C 19	135.6	3.5	6866	24	AB152666	Human immune syste
C 20	131.8	3.4	574	24	ABQ21886	Oligonucleotide fo
C 21	131.8	3.4	574	24	ABQ21887	Oligonucleotide fo
C 22	131.8	3.4	574	24	ABQ32920	Oligonucleotide fo
C 23	131.8	3.4	574	24	ABQ32921	Oligonucleotide fo
C 24	131.8	3.4	6866	24	AB149320	Human polynucleoti
C 25	131.8	3.4	6866	24	AB152667	Human immune syste
C 26	112.4	2.9	589	21	AA167599	Fusarium venenatum
C 27	111.4	2.9	842	23	ABV03504	Human prostate exp
C 28	85.4	2.2	405	21	ABO55897	Human colon cancer
C 29	80	2.1	430	24	ABO55897	Human ovarian anti
C 30	79.4	2.1	570	23	ABV22282	Human prostate exp
C 31	79.4	2.1	570	23	ABV28116	Human prostate exp
C 32	79.4	2.1	605	23	ABV33803	Human prostate exp
C 33	79.4	2.1	605	23	ABV42696	Human prostate exp
C 34	79.4	2.1	871	23	ABV12673	Human prostate exp
C 35	72.2	1.9	288	20	AAV89510	EST clone COL151.
C 36	72.2	1.9	110	21	AA12928	Human secreted pro
C 37	70.6	1.8	318	22	AB16642	Human nervous syst
C 38	70.6	1.8	6318	22	AA199118	Human excretory re
C 39	70.6	1.8	6318	22	AA163468	Human kidney relat
C 40	70.6	1.8	32127	22	AA199255	Human excretory re
C 41	70.6	1.8	32127	22	AA163605	Human kidney relat
C 42	70.6	1.8	72215	22	AAK66832	Human immune/haema
C 43	70.6	1.8	236303	22	AA511614	Human genomic DNA
C 44	70.2	1.8	9566	22	ABR20411	Human nervous syst
C 45	70.2	1.8	9566	22	AA157121	Human musculoskele

ALIGNMENTS

RESULT 1
ABK35494
ID ABK35494 standard; DNA: 3857 BP.
XX
AC ABK35494:
XX
DT 08-MAY-2002 (first entry)
XX
DE Human endometrial cancer related gene, ERCC4.
XX
KW Human; ds: gene: endometrial cancer; differential expression;
KW DNA microarray; protein microarray.
XX
OS Homo sapiens.
XX
PN W0200209573-A2.
XX
PD 07-FEB-2002.
XX
PF 31-JUL-2001; 2001WO-US24104.
XX
PR 31-JUL-2000; 2000US-221735P.
XX
PA (BGHM) BRIGHAM & WOMEN'S HOSPITAL INC.
XX
PI Mutter GL;
XX
DR WPI: 2002-179967/23.
XX
PT P-PSDB: AAU84274.
PT
PT Diagnosing endometrial cancer comprises determining expression of
nucleic acid molecules or expression products that are differentially

PT expressed in normal and malignant endometrium -
 XX Claim 1: Page 64-66; 233pp; English.
 XX The invention relates to diagnosing endometrial cancer in a subject
 CC suspected of having endometrial cancer comprising determining the
 CC expression of a set of nucleic acid molecules or expression products in
 CC an endometrial sample suspected of being cancerous, where the set of
 CC nucleic acid molecules comprises at least 2 nucleic acid molecules
 CC selected from 50 fully defined sequences as given in the specification.
 CC The nucleic acids are used as an array of at least 2 of the 50
 CC nucleic acids bound to a solid substrate. Also included is a solid-phase
 CC protein microarray comprising at least 2 antibodies or its antigen
 CC binding fragments, that specifically bind at least 2 different
 CC polypeptides from the 50 fully defined sequences as given in the
 CC specification, fixed to a solid substrate. The methods and arrays are
 CC useful for the diagnosis of endometrial cancer, selecting and monitoring
 CC treatment regimes and identification of lead compounds useful for the
 CC treatment of endometrial cancer. The present sequence is one of 50
 CC genes differentially expressed between cancerous and non-cancerous
 CC samples.

XX Sequence 3857 BP; 1134 A; 851 C; 850 G; 1022 T; 0 other:

Query Match 100.0%; Score 3857; DB 24; Length 3857;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 3857; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGAAGACCTTCCATGAGTACAGGCGCGCTCGACGCATTGCCGCGCTGCTG 60
 DB 1 GGAAGACCTTCCATGAGTACAGGCGCGCTCGACGCATTGCCGCGCTGCTG 60
 QY 61 GACTACAGACGACAGCTGCTGCTGCAACCTGCTGACACTGACGAGGCTAGTGTGCGCC 120
 DB 61 GACTACAGACGACAGCTGCTGCTGCAACCTGCTGACACTGACGAGGCTAGTGTGCGCC 120
 QY 121 CGGCGGCTGCGCGCGGACCGGCTCTCTACCACTTCTCCAGCTGACCTCCACCGACGC 180
 DB 121 CGGCGGCTGCGCGCGGACCGGCTCTCTACCACTTCTCCAGCTGACCTCCACCGACGC 180
 QY 181 TGGCTGTGCTGTGCTGCTGCAACCGCAGCGCGGAGAGAGATTTTATCATCAGCTG 240
 DB 181 TGGCTGTGCTGTGCTGCTGCAACCGCAGCGCGGAGAGATTTTATCATCAGCTG 240
 QY 241 AAGATGAAAGAGATTAACACCTCCCTCGCGCTGTAACAATGAATCAACAACAGT 300
 DB 241 AAGATGAAAGAGATTAACACCTCCCTCGCGCTGTAACAATGAATCAACAACAGT 300
 QY 301 CGCTATGAAGTTTACACACAGGTGCTTATATTTGCGACAGTAGATCTTGTGTT 360
 DB 301 CGCTATGAAGTTTACACACAGGTGCTTATATTTGCGACAGTAGATCTTGTGTT 360
 QY 361 GACTTGTGACGTGATGAACTTACCTGAGATTTATCTGCGATCTTGGTATTAAGCC 420
 DB 361 GACTTGTGACGTGATGAACTTACCTGAGATTTATCTGCGATCTTGGTATTAAGCC 420
 QY 421 CACAGATATATGAGTCTGTCAAGAAACATTCATCTTGGCGCTCTTGGCCGAAAC 480
 DB 421 CACAGATATATGAGTCTGTCAAGAAACATTCATCTTGGCGCTCTTGGCCGAAAC 480
 QY 481 AAACGTGTTTTATTAAGCTTACACAGCAATGCTGTGCTTGTGATGACTGTTTTGT 540
 DB 481 AAACGTGTTTTATTAAGCTTACACAGCAATGCTGTGCTTGTGATGACTGTTTTGT 540
 QY 541 CATGTGAAAGAGTGAATGAAATCTTTTGTGAGAACTGATCTGTGCGCAAGCTTC 600
 DB 541 CATGTGAAAGAGTGAATGAAATCTTTTGTGAGAACTGATCTGTGCGCAAGCTTC 600
 QY 601 CATGTGAGCACTAATCTATTTTGAAGACAGCAAACTGAACTGTAGAAATCCATGTT 660
 DB 601 CATGTGAGCACTAATCTATTTTGAAGACAGCAAACTGAACTGTAGAAATCCATGTT 660
 QY 661 TCTATGACACCTACATGCTGCTATACAGACTGCTATGACATTTTAAATGATGT 720

DB 661 TCTATGACACCTACATGCTGCTATACAGACTGCTATACATTTTAAATGATGT 720
 QY 721 CTAAGAGAACTAAATGCTTACCATGCTGAAAGTGAAGATTTATCTTGAAGAAAT 780
 DB 721 CTAAGAGAACTAAATGCTTACCATGCTGAAAGTGAAGATTTATCTTGAAGAAAT 780
 QY 781 GCATTTGAAACCTTTTGAAGACAAATCCGCTATATCTGATCTTGTGCGACAG 840
 DB 781 GCATTTGAAACCTTTTGAAGACAAATCCGCTATATCTGATCTTGTGCGACAG 840
 QY 841 CTTGAGCCAAAGCTAATCTTATGTTCAAGATTTGAAGATTTACGACTTGTGAC 900
 DB 841 CTTGAGCCAAAGCTAATCTTATGTTCAAGATTTGAAGATTTACGACTTGTGAC 900
 QY 901 TATCTCTCTGATGATGATGCTGACATTTCTTAATCTTCTGGAATCTCTGAAGCAAG 960
 DB 901 TATCTCTCTGATGATGATGCTGACATTTCTTAATCTTCTGGAATCTCTGAAGCAAG 960
 QY 961 GAAAAAGCTTTGCTGCAATTCAGGTGCGCTTCTTCTGACCTCCAGCACTGATGTTT 1020
 DB 961 GAAAAAGCTTTGCTGCAATTCAGGTGCGCTTCTTCTGACCTCCAGCACTGATGTTT 1020
 QY 1021 ATAAATGCTCGACCAAGGTTTATCATCTTCCAGATGCCAAATGAGTAAAGAAAA 1080
 DB 1021 ATAAATGCTCGACCAAGGTTTATCATCTTCCAGATGCCAAATGAGTAAAGAAAA 1080
 QY 1081 ATATCTGAAAAAATGGAATTTAAGAAAGGGAACAAAAAGAACTGCTCTAGAA 1140
 DB 1081 ATATCTGAAAAAATGGAATTTAAGAAAGGGAACAAAAAGAACTGCTCTAGAA 1140
 QY 1141 AGCAACCCAAAGTGGAGGACACTGACTGAATTTTAAAGAAATTAAGCAGAAAAATAG 1200
 DB 1141 AGCAACCCAAAGTGGAGGACACTGACTGAATTTTAAAGAAATTAAGCAGAAAAATAG 1200
 QY 1201 GAGAGTGAAGCTCTTGTGTGCTCAGGTCAATCTGATTTTGTGCAAGTGTGACCAAGCA 1260
 DB 1201 GAGAGTGAAGCTCTTGTGTGCTCAGGTCAATCTGATTTTGTGCAAGTGTGACCAAGCA 1260
 QY 1261 TGTTCCTCAGCTGAGACATATATCATCTTGTGAGCGGAGGCTTCTTATTTGAGGCTCTAC 1320
 DB 1261 TGTTCCTCAGCTGAGACATATATCATCTTGTGAGCGGAGGCTTCTTATTTGAGGCTCTAC 1320
 QY 1321 AGGAAACCTTTGAGAGAGATAGCAAAAGCTGAAGATCTGATGAATTTAGAGAGAA 1380
 DB 1321 AGGAAACCTTTGAGAGAGATAGCAAAAGCTGAAGATCTGATGAATTTAGAGAGAA 1380
 QY 1381 GACAGTTCAAGAGATTTAGGAATCTCACAAAAGACTTAAGACCCCAAAACAAAGAA 1440
 DB 1381 GACAGTTCAAGAGATTTAGGAATCTCACAAAAGACTTAAGACCCCAAAACAAAGAA 1440
 QY 1441 CGGCGCTTCTACCAAGAAAGAACTCTCAAAAGAAAAAGGAACTTGAACCTTAACCTCAA 1500
 DB 1441 CGGCGCTTCTACCAAGAAAGAACTCTCAAAAGAAAAAGGAACTTGAACCTTAACCTCAA 1500
 QY 1501 ATGTTAGGAAACCTGAGAACTGGAAGAGAGAGATGTCGAGCAAGATATCTGCTCA 1560
 DB 1501 ATGTTAGGAAACCTGAGAACTGGAAGAGAGAGATGTCGAGCAAGATATCTGCTCA 1560
 QY 1561 GAAATTAAGCAGTAGCCAGAAAGCTCCCGGAGAAATTAAGCATAGCAATTTGATGTA 1620
 DB 1561 GAAATTAAGCAGTAGCCAGAAAGCTCCCGGAGAAATTAAGCATAGCAATTTGATGTA 1620
 QY 1621 AATTTGCAATCGATGCTGTTTGGAAATCTGAAAGAACCCCTCACTATCATCCATCCG 1680
 DB 1621 AATTTGCAATCGATGCTGTTTGGAAATCTGAAAGAACCCCTCACTATCATCCATCCG 1680
 QY 1681 CTTCTGCTGCTGAGCAAGCCCTATGCTGACAAAGGCTACTACATGAAGTGAAGCCCAAGA 1740
 DB 1681 CTTCTGCTGCTGAGCAAGCCCTATGCTGACAAAGGCTACTACATGAAGTGAAGCCCAAGA 1740
 QY 1741 TACGTGCTCTTTTATGACGACAGCTAACCTTTGTTGGCAGCTTCAAAATTTACAGCGCG 1800

D 1741 TACGTGGTCTTTATGACGACGACCTAACCTTTGTGGCAGCTTGAAATTTACAGGGCG 1800
Q 1801 AGTAGGCTGGGAACCTCTGAGGGTTACTTTCTTATATACGAGGTTCAACGAGAA 1860
D 1801 AGTAGGCTGGGAACCTCTGAGGGTTACTTTCTTATATACGAGGTTCAACGAGAA 1860
Q 1861 CAAAGCATATCAGTGTGCTTGGGAAAGAAAGAAAGCTTTTGAAGAACTATAGGAA 1920
D 1861 CAAAGCATATCAGTGTGCTTGGGAAAGAAAGAAAGCTTTTGAAGAACTATAGGAA 1920
Q 1921 AAAGCAAGCATGTTGTCCTGGAAGAAAGAGAGAGAGATGAACAACTTAGACTTA 1980
D 1921 AAAGCAAGCATGTTGTCCTGGAAGAAAGAGAGAGATGAACAACTTAGACTTA 1980
Q 1981 GTAGAGGACAGATCGCATGATGTTTCCAGTACACTGGAAGGCGGGGCGCAGAA 2040
D 1981 GTAGAGGACAGATCGCATGATGTTTCCAGTACACTGGAAGGCGGGGCGCAGAA 2040
Q 2041 CAGAATGGTACACAGCAAAAGCATAGTGTGATATGCGTGAATTTGCAAGTGAAGTTCCA 2100
D 2041 CAGAATGGTACACAGCAAAAGCATAGTGTGATATGCGTGAATTTGCAAGTGAAGTTCCA 2100
Q 2101 TCTCTGATCCATGCTGGGGGATTTGACATTTGAACCCGTAAGTTAGAGTTGAGATTAC 2160
D 2101 TCTCTGATCCATGCTGGGGGATTTGACATTTGAACCCGTAAGTTAGAGTTGAGATTAC 2160
Q 2161 ATCTCATCTCAGAAATGTGGTGAGGCGCAAGAGTATCAGTGAATTTAATGGGCTTTTA 2220
D 2161 ATCTCATCTCAGAAATGTGGTGAGGCGCAAGAGTATCAGTGAATTTAATGGGCTTTTA 2220
Q 2221 AATAAGGCGCCCTCTACAGCCAGTGTGATCTCATGTCCTCCCTACTACAGCCGTCCTG 2280
D 2221 AATAAGGCGCCCTCTACAGCCAGTGTGATCTCATGTCCTCCCTACTACAGCCGTCCTG 2280
Q 2281 CTCTGATGATGTTGACCCCTAGCAAGCCCTTTCTCTCACTTCCCGAGGTCCTGTTT 2340
D 2281 CTCTGATGATGTTGACCCCTAGCAAGCCCTTTCTCTCACTTCCCGAGGTCCTGTTT 2340
Q 2341 CAGAGATCTCCAGCAATGACATTAAGTTCAGAACTCACTCTTCTTACACTTCCCTCC 2400
D 2341 CAGAGATCTCCAGCAATGACATTAAGTTCAGAACTCACTCTTCTTACACTTCCCTCC 2400
Q 2401 AGACTAGGATTTCTGTGTCGCCCTCTCTCATGCAAGGCGGAGTTTGGAGAGCTG 2460
D 2401 AGACTAGGATTTCTGTGTCGCCCTCTCTCATGCAAGGCGGAGTTTGGAGAGCTG 2460
Q 2461 AAACAAAGCAAGCCACAGCCGATGCGGCGACAGCACTGCGCATTTACAGCATTTCTGA 2520
D 2461 AAACAAAGCAAGCCACAGCCGATGCGGCGACAGCACTGCGCATTTACAGCATTTCTGA 2520
Q 2521 ACCCTCCCGAGTACAGAAATATATGCTGTGCCCAAGCTCTTGTAAATATGCCA 2580
D 2521 ACCCTCCCGAGTACAGAAATATATGCTGTGCCCAAGCTCTTGTAAATATGCCA 2580
Q 2581 GGGGTGATGCAAAAACCTGCGCTCTTGTGACACACAGTGAAGACATGCGAGAAATTA 2640
D 2581 GGGGTGATGCAAAAACCTGCGCTCTTGTGACACACAGTGAAGACATGCGAGAAATTA 2640
Q 2641 GCAGCCCTGTACAGAGAGAGCTCAGAGTATTTCTGGGGAATGCTGCAAAATGCCAAACAG 2700
D 2641 GCAGCCCTGTACAGAGAGAGCTCAGAGTATTTCTGGGGAATGCTGCAAAATGCCAAACAG 2700
Q 2701 CTTATATGATTTTCAATTCACACTCTTTTGCAGAAAGTCAAAAAGGAGGAGAAAG 2760
D 2701 CTTATATGATTTTCAATTCACACTCTTTTGCAGAAAGTCAAAAAGGAGGAGAAAG 2760
Q 2761 TGAACAGTATGAGGCTGTTTCTTATCCAGTGTGATTTTCAAGGCGCTCCTTCCGAGA 2820
D 2761 TGAACAGTATGAGGCTGTTTCTTATCCAGTGTGATTTTCAAGGCGCTCCTTCCGAGA 2820
Q 2821 CATCATAGTATTAATTAATGTTGTTGCTATTTCACTTCTTCCAAATGCTCTTAATG 2880
D 2821 CATCATAGTATTAATTAATGTTGTTGCTATTTCACTTCTTCCAAATGCTCTTAATG 2880

Q 2881 ATTGACGGTGGACAGAGAGCCAGGATTCCTCTGAACTCTGCAAGTATGACATCACTTG 2940
D 2881 ATTGACGGTGGACAGAGAGCCAGGATTCCTCTGAACTCTGCAAGTATGACATCACTTG 2940
Q 2941 AACTGTGCTGTGCTGCTCTTTTCTCCCTGACAGCCGCTATGCGGCTTACGATGTTT 3000
D 2941 AACTGTGCTGTGCTGCTCTTTTCTCCCTGACAGCCGCTATGCGGCTTACGATGTTT 3000
Q 3001 CTTTAAATAGAGTTTGTACAGATCAGGTAAAGTTCTTACAAAGTATACAGAAAGTATG 3060
D 3001 CTTTAAATAGAGTTTGTGTACAGATCAGGTAAAGTTCTTACAAAGTATACAGAAAGTATG 3060
Q 3061 AAACCTTCTGATCTTCAAGATCTCATTTAGAAAGAAATATGCTAAGCCTGGCATGGA 3120
D 3061 AAACCTTCTGATCTTCAAGATCTCATTTAGAAAGAAATATGCTAAGCCTGGCATGGA 3120
Q 3121 CGGTGACAGGAGGAGAAAGACAGGACACAGAAAGCTACATTTTAAACAGTCTTGTTA 3180
D 3121 CGGTGACAGGAGGAGAAAGACAGGACACAGAAAGCTACATTTTAAACAGTCTTGTTA 3180
Q 3181 TCTAGTGAACATTAATTAACAGTCTTAATGCACTTATACCATTTCTGCTGGCTTCCA 3240
D 3181 TCTAGTGAACATTAATTAACAGTCTTAATGCACTTATACCATTTCTGCTGGCTTCCA 3240
Q 3241 AATCGGCTTTGCTGCTGTGTGCTGCTGAGAGCTTGAACATGATGTTGTAGGAAATCA 3300
D 3241 AATCGGCTTTGCTGCTGTGTGCTGCTGAGAGCTTGAACATGATGTTGTAGGAAATCA 3300
Q 3301 TGTCTGACCCCTTGTGTCTACAAAGAGCCCTCTGGAACCTGAGAAAGAAATCTCTTG 3360
D 3301 TGTCTGACCCCTTGTGTCTACAAAGAGCCCTCTGGAACCTGAGAAAGAAATCTCTTG 3360
Q 3361 CCATTCCTGACAGTGTCTCTACACATTTTCTTCACTGCTCATPACTTCTGCTGTGCG 3420
D 3361 CCATTCCTGACAGTGTCTCTCTACACATTTTCTTCACTGCTCATPACTTCTGCTGTGCG 3420
Q 3421 TCTAAGAAATTTTCAATGAGGCTTCTCTACTAATTTGAAGACAGTCTCTCAAAAATG 3480
D 3421 TCTAAGAAATTTTCAATGAGGCTTCTCTACTAATTTGAAGACAGTCTCTCTCAAAAATG 3480
Q 3481 GTTGACATGCTCTTAATGACCTTACATATGATAGCATATATTAATTTGATGTTCCA 3540
D 3481 GTTGACATGCTCTTAATGACCTTACATATGATAGCATATATTAATTTGATGTTCCA 3540
Q 3541 AATTAGTATTTTAAAGCAAAATGATTAACGTTTGGAAAAAGTAAATGATGAAGAGCT 3600
D 3541 AATTAGTATTTTAAAGCAAAATGATTAACGTTTGGAAAAAGTAAATGATGAAGAGCT 3600
Q 3601 CTAGAAATTCGAATTTTGGACATATTCAGTCTCTTAATTAACAGATCCCTAAGTCCA 3660
D 3601 CTAGAAATTCGAATTTTGGACATATTCAGTCTCTTAATTAACAGATCCCTAAGTCCA 3660
Q 3661 GCTGGCTAGTTACAGAGTTTTTTCAGACTCTCTGTTTCTCAGCTTATATCTTAAGAC 3720
D 3661 GCTGGCTAGTTACAGAGTTTTTTCAGACTCTCTGTTTCTCAGCTTATATCTTAAGAC 3720
Q 3721 ACCAGCATATATCTCTAGAAATTAACAACCTTAATGSCAGTGAAGCGGAGATGCCACT 3780
D 3721 ACCAGCATATATCTCTAGAAATTAACAACCTTAATGSCAGTGAAGCGGAGATGCCACT 3780
Q 3781 GCACCCCTGCGTGGGCGACAGAGTGAAGCTTGTCTCTATTAACAAGAAAAAGAAAGA 3840
D 3781 GCACCCCTGCGTGGGCGACAGAGTGAAGCTTGTCTCTATTAACAAGAAAAAGAAAGA 3840
Q 3841 AATCAACCTTAAGTCA 3857
D 3841 AATCAACCTTAAGTCA 3857

RESULT 2
AAT97897
ID AAT97897 standard; cdNA: 2890 BP.

XX AAT97897;
AC 11-MAY-1998 (first entry)
XX
DE Human DNA repair enzyme RAD cDNA.
XX
KM RAD: DNA repair enzyme; human; xeroderma pigmentosum; colon cancer;
hereditary nonpolyposis coli; HNPPC; diagnosis; gene therapy; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 25..2742
FT /*tag= a
XX
XX WO9742209-A1.
XX
XX 13-NOV-1997.
XX
XX 03-MAY-1996; 96MO-US06221.
XX
XX 03-MAY-1996; 96MO-US06221.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
PA (ICRF-) ICRF HALL LAB CLARE.
PA (UYRO-) UNIV ROTTERDAM ERASMUS.
XX
PI Ariza RR, Hoelmakers JHJ, Ruben SM, Sijbers A;
PI Wei YF, Wood RD;
XX
XX WPI: 1997-558900/51.
XX P-PSDB: AAM36879.
XX
XX Polynucleotide encoding human DNA repair enzyme, RAD - useful to
PT develop products to diagnose and treat, e.g. xeroderma pigmentosum
PT and colon cancer
XX
XX Claim 4: Page 60-61; 81pp; English.
XX
XX This cDNA sequence includes a coding region for the human DNA
CC repair enzyme RAD (see AAM36879). Degenerate primers based on the
CC homology between Saccharomyces cerevisiae Rad1, Schizosaccharomyces
CC pombe Rad16+ and Drosophila melanogaster mei-9 gene products were
CC used in a RT-PCR with total HeLa cell RNA to amplify a fragment
CC encoding the homologous human gene. The PCR product was used as a
CC probe to identify a positive clone from a human testis cDNA
CC library. The sequences were then used to obtain the full-length
CC human RAD cDNA. The isolated polynucleotide can be used for the
CC recombinant production of RAD polypeptides in host cells. Methods
CC are disclosed for assessing RAD expression in cells by determining
CC RAD polypeptides or RAD-encoding mRNA, for treating diseases
CC associated with defective DNA repair such as xeroderma pigmentosum,
CC colon cancer and hereditary nonpolyposis coli (HNPPC) in vitro, ex
CC vivo or in vivo by exposing cells to RAD polypeptides or
CC polynucleotides, for assaying genetic variation and aberrations in
CC RAD genes, and for administering a RAD polypeptide or polynucleotide
CC to an organism to augment RAD function or remediate RAD dysfunction.
XX
XX Sequence 2890 BP; 862 A; 630 C; 672 G; 726 T; 0 other:
SO
Query Match 74.5%; Score 2875; DB 18; Length 2890;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 2878; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 142 CTCCTACCACTTTCTCCAGCTGCACGACCCAGCCTGCCTGCTGCTCAAC 201
DB 121 CTCCTACCACTTTCTCCAGCTGCACGACCCAGCCTGCCTGCTGCTCAAC 180
QY 202 ACGAGCCGCGGAGAGAGATATTTATCAATCACTGAGATGAGAGAGTTGAAC 261
DB 181 ACGAGCCGCGGAGAGAGATATTTATCAATCACTGAGATGAGAGAGTTGAAC 240
QY 262 CTCCTCCGCGGCTGTAACAAATGAAATCAACACAGCTGCTGATGACACAA 321
DB 241 CTCCTCCGCGGCTGTAACAAATGAAATCAACACAGCTGCTGATGACACAA 300
QY 322 GGTGCTTATATTTGCGACAGATGATGATGATGATGATGATGATGATGAT 381
DB 301 GGTGCTTATATTTGCGACAGATGATGATGATGATGATGATGATGATGAT 360
QY 382 CCTTCAGATTTAATTAATGATGATGATGATGATGATGATGATGATGAT 441
DB 361 CCTTCAGATTTAATTAATGATGATGATGATGATGATGATGATGATGAT 420
QY 442 CAAGAAGCATTTGATCTGGGCTCTTGGCCAGAAAACAAACGCTGTTTAAAGCT 501
DB 421 CAAGAAGCATTTGATCTGGGCTCTTGGCCAGAAAACAAACGCTGTTTAAAGCT 480
QY 502 TTCACAGCAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 561
DB 481 TTCACAGCAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540
QY 562 AATCTTTTGTGAGAACTGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 621
DB 541 AATCTTTTGTGAGAACTGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 600
QY 622 TTGAAACAGCAAAACCTGAGATGAGAAATGCAATGCTGCTGCTGCTGCT 681
DB 601 TTGAAACAGCAAAACCTGAGATGAGAAATGCAATGCTGCTGCTGCTGCT 660
QY 682 GCTATACAGATGCTATGATGATGATGATGATGATGATGATGATGATGAT 741
DB 661 GCTATACAGATGCTATGATGATGATGATGATGATGATGATGATGATGAT 720
QY 742 AACCCATGCTTGAAGTGAAGATTTATCTTGAAGAAATGCTATGAGAACTTTGAC 801
DB 721 AACCCATGCTTGAAGTGAAGATTTATCTTGAAGAAATGCTATGAGAACTTTGAC 780
QY 802 AAGACAAATCCGCAATATGATGATGATGATGATGATGATGATGATGATGAT 861
DB 781 AAGACAAATCCGCAATATGATGATGATGATGATGATGATGATGATGATGAT 840
QY 862 TTAGTTCAAGATTTGAAGATTTAGCAATTTGCTGCTGCTGCTGCTGCTGCT 921
DB 841 TTAGTTCAAGATTTGAAGATTTAGCAATTTGCTGCTGCTGCTGCTGCTGCT 900
QY 922 GTCACATTTCTTAATTTGATTTGATTTGATTTGATTTGATTTGATTTGAT 981
DB 901 GTCACATTTCTTAATTTGATTTGATTTGATTTGATTTGATTTGATTTGAT 960
QY 982 TCAGTTGCTGCTTTCTTACTCCACACCTGATGTTATTAATGCTGAGCAAGGTT 1041
DB 961 TCAGTTGCTGCTTTCTTACTCCACACCTGATGTTATTAATGCTGAGCAAGGTT 1020
QY 1042 TATCATCTTCCAGATGCCAAATGAGTAAAGAAAGAAATTTCTGAAAGTGAAT 1101
DB 1021 TATCATCTTCCAGATGCCAAATGAGTAAAGAAAGAAATTTCTGAAAGTGAAT 1080
QY 1102 AAAGAAGGGAAGAAACAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA 1161
DB 1081 AAAGAAGGGAAGAAACAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA 1140
QY 1162 CTGACTGAAGTATTAAGAAATTTAGGAGAAATTAAGAGAGTGAAGCTTTGTTGCT 1221
DB 1141 CTGACTGAAGTATTAAGAAATTTAGGAGAAATTAAGAGAGTGAAGCTTTGTTGCT 1200
QY 1222 CCAAGTCAAGTACTGATTTTGTGCAAGTATGACCCGAACATGTTCCAGCTGACAGACTAT 1281

Db 1201 CCAGGTCAGACTGATTTGTGCAAGTGAAGCCGAACATGTTCCAGCTGAGAGACTAT 1260
 QY 1282 ATCACTCTTGGAGCGAGGCTCTTCTATTGAGGCTCTACAGGAAACCTTTGAGAGAT 1341
 Db 1261 ATCACTCTTGGAGCGAGGCTCTTCTATTGAGGCTCTACAGGAAACCTTTGAGAGAT 1320
 QY 1342 AGCAAAAGCTGAAGAGTCTGAGTGAATTTAGAGAGAAACAGTTCAAAGAGATTAAG 1401
 Db 1321 ACCAAAGCTGAAGAGTCTGAGTGAATTTAGAGAGAAACAGTTCAAAGAGATTAAG 1380
 QY 1402 AATCTCACAAAAGACCTTAAGACCCCAAAACAAAGAACGGGCTCTCCAAAGAAAGA 1461
 Db 1381 AATCTCACAAAAGACCTTAAGACCCCAAAACAAAGAACGGGCTCTCCAAAGAAAGA 1440
 QY 1462 ACCCTCAAAAAGAAAAAGGAAAGTTGACCTTAACCTCAATGGTAGAAACCTGAAGA 1521
 Db 1441 ACCCTCAAAAAGAAAAAGGAAAGTTGACCTTAACCTCAATGGTAGAAACCTGAAGA 1500
 QY 1522 CTGGAAGAGGAAGAGATGTCGAGAGAGATGTCGAGAAATTAAGCAGTACCCAGAA 1581
 Db 1501 CTGGAAGAGGAAGAGATGTCGAGAGAGATGTCGAGAAATTAAGCAGTACCCAGAA 1560
 QY 1582 AGCTGCCGGAAGAAATTAAGCATGAAGATTTGATGAATTTTCATCCGATGCTGCT 1641
 Db 1561 AGCTGCCGGAAGAAATTAAGCATGAAGATTTGATGAATTTTCATCCGATGCTGCT 1620
 QY 1642 TTCGGAATCTGAAGAAACCCCTCACTATCATCCGCTCTGAGGTTGACAGCAGCC 1701
 Db 1621 TTCGGAATCTGAAGAAACCCCTCACTATCATCCGCTCTGAGGTTGACAGCAGCC 1680
 QY 1702 TATGCTCTGACAAAGGCTACTACATGAAGTGAGCCAAAGTACGTGGTTCTTTATGACGA 1761
 Db 1681 TATGCTCTGACAAAGGCTACTACATGAAGTGAGCCAAAGTACGTGGTTCTTTATGACGA 1740
 QY 1762 GAGCTACCTTTGTCGAGGCTGGAATTTACAGGGCCAGTAGGCTGAGAAACCTG 1821
 Db 1741 GAGCTACCTTTGTCGAGGCTGGAATTTACAGGGCCAGTAGGCTGAGAAACCTG 1800
 QY 1822 AAGGTTTACTTTCTTATATACGAGGTTCACTGAGAACAAAGCTATCTCACTGTTG 1881
 Db 1801 AAGGTTTACTTTCTTATATACGAGGTTCACTGAGAACAAAGCTATCTCACTGTTG 1860
 QY 1882 CGGAAAGAAAGAGAGCTTTTGAAAACTCATAGGGAAAAAGCAAGCATGTTGCC 1941
 Db 1861 CGGAAAGAAAGAGAGCTTTTGAAAACTCATAGGGAAAAAGCAAGCATGTTGCC 1920
 QY 1942 GAAGAAAGAGAGAGAGATGAACAACTTAGACCTAGTAAGAGGACAGCATCTGCA 2001
 Db 1921 GAAGAAAGAGAGAGAGATGAACAACTTAGACCTAGTAAGAGGACAGCATCTGCA 1980
 QY 2002 GATGTTTCCACTGACACTGGAAGCCGGTGGCAGAGACAGAAATGTCACAGCAAGC 2061
 Db 1981 GATGTTTCCACTGACACTGGAAGCCGGTGGCAGAGACAGAAATGTCACAGCAAGC 2040
 QY 2062 ATAGTTTGATATGCGTGAATTTGCAAGTGAGCTTCATCTCTGATCCATCGTCGGG 2121
 Db 2041 ATAGTTTGATATGCGTGAATTTGCAAGTGAGCTTCATCTCTGATCCATCGTCGGG 2100
 QY 2122 ATTGACATTGAACCCGTGACTTTAGAGTTGAGATTAATCTCTCACTCAGAAATGTGC 2181
 Db 2101 ATTGACATTGAACCCGTGACTTTAGAGTTGAGATTAATCTCTCACTCAGAAATGTGC 2160
 QY 2182 GTGAGAGCGAAGAGTATCAGTATTTAATCGGCTCTTAAATACGGCCGCTCTACAG 2241
 Db 2161 GTGAGAGCGAAGAGTATCAGTATTTAATCGGCTCTTAAATACGGCCGCTCTACAG 2220
 QY 2242 CAGTGATCTCATGTCGCGCTACTACAAGCGTCCCGTCTCTGATTGAGTTGACCC 2301
 Db 2221 CAGTGATCTCATGTCGCGCTACTACAAGCGTCCCGTCTCTGATTGAGTTGACCC 2280
 QY 2302 AGCAAGCCTTCTCTCTACTCCGAGGTGCTTCTCAGAGATCTCCAGCATATGC 2361

Db 2281 AGCAAGCCTTCTCTCTACTCTCCAGAGTGCCCTTTGACAGAGATCTCCAGCAATGAC 2340
 QY 2362 ATAGTTTCAAACTCACTCTTCTTACACTTCACTTCCAGACTACAGATTCGTGGTC 2421
 Db 2341 ATAGTTTCAAACTCACTCTTCTTACACTTCACTTCCAGACTACAGATTCGTGGTC 2400
 QY 2422 CCTCTCTCATGCAAGCGCGGAGTTGTTGAGAGACTGAAACAAAGCAAGCCACAGCT 2481
 Db 2401 CCTCTCTCATGCAAGCGCGGAGTTGTTGAGAGACTGAAACAAAGCAAGCCACAGCT 2460
 QY 2482 GATGGGCGCAGACACTGGCCATTACAGAGATTCGAAACCTTCCGAGTCAGAGAG 2541
 Db 2461 GATGGGCGCAGACACTGGCCATTACAGAGATTCGAAACCTTCCGAGTCAGAGAG 2520
 QY 2542 TATATCTGTCGTCGCCAAGACTCTTGTAAAAATGCCAGGGTGAATGCCAAAACTGC 2601
 Db 2521 TATATCTGTCGTCGCCAAGACTCTTGTAAAAATGCCAGGGTGAATGCCAAAACTGC 2580
 QY 2602 CGCTCTTATGACACACGTTAAGAACATGCGAGAAATTAAGCAGCCCTGTACAGAGAG 2661
 Db 2581 CGCTCTTATGACACACGTTAAGAACATGCGAGAAATTAAGCAGCCCTGTACAGAGAG 2640
 QY 2662 CTCACAGATATCTGGGGAAATGCTGCAATGCCAAACACTTTATGATTTCAATTCACACC 2721
 Db 2641 CTCACAGATATCTGGGGAAATGCTGCAATGCCAAACACTTTATGATTTCAATTCACACC 2700
 QY 2722 TCTTTTGCAAGATCTATCAAAAAGGAAAAAGGAAAGTGAACAGTGAATGGCTGTTTC 2781
 Db 2701 TCTTTTGCAAGATCTATCAAAAAGGAAAAAGGAAAGTGAACAGTGAATGGCTGTTTC 2760
 QY 2782 TATCCAGTCCCTGTACTTTTCAAGCGCTCTTGCAGACATCATAGTCAATTAATTT 2841
 Db 2761 TATCCAGTCCCTGTACTTTTCAAGCGCTCTTGCAGACATCATAGTCAATTAATTT 2820
 QY 2842 ATTTGTTCTATTTATCTTTTCCATGCTTTATATATTTGACGGTGACAGAAC 2901
 Db 2821 ATTTGTTCTATTTATCTTTTCCATGCTTTATATATTTGACGGTGACAGAGTT 2880
 QY 2902 CAG 2904
 Db 2881 CAG 2883

RESULT 3
 AAF16235
 ID AAF16235 standard: cDNA: 2900 BP.
 XX AAF16235;
 AC
 DT 13-MAR-2001 (first entry)
 XX
 DE Human prostate cancer antigen nucleotide sequence SEQ ID NO:670.
 XX
 OS Homo sapiens.
 PN WO20005174-A1.
 XX
 PD 21-SEP-2000.
 XX
 PF 08-MAR-2000; 2000WO-US05988.
 XX
 PR 12-MAR-1999; 99US-0124270.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX (ROSE/) ROSEN C A.
 XX

Human; prostate cancer; prostate cancer antigen: detection; diagnosis;
 neuroprotective; cytosolic; cardioactive; immunomodulatory; muscular;
 vulnery; gastrointestinal; nephrotoxic; antiinfective; gynaecological;
 antibacterial; gene therapy; neural; immune; reproductive; renal;
 gastrointestinal; pulmonary; cardiovascular; proliferative disorder;
 wound; infectious disease; ss.

PI Rosen CA, Ruben SM:
XX WPI: 2000-587513/55.
DR P-PSDB; AAB57032.
XX
PT Prostate cancer associated gene sequences, referred to as prostate
PT cancer antigens, useful for treatment, prevention, and diagnosis of
PT disorders such as prostate cancer -
XX
PS Claim 1; Page 1106-1107; 2338bp; English.
XX
CC AAF15566 to AAF16505 encode the human prostate cancer associated
CC proteins, called prostate cancer antigens, given in AAB56363 to AAB57302.
CC The prostate cancer antigens can have neuroprotective, cytostatic,
CC cardioactive, immunomodulatory, muscular, vulnery, gastrointestinal,
CC nephrotoxic, antiinfective, gynecological and antibacterial activities,
CC and can be used in gene therapy. The prostate cancer antigen
CC polynucleotides may be used for detection of prostate cancer, chromosome
CC identification, as chromosome markers, and for numerous other diagnostic
CC or research purposes. The prostate cancer antigens may be used to treat
CC disorders such as neural, immune, muscular, reproductive,
CC gastrointestinal, pulmonary, cardiovascular, renal, and proliferative
CC disorders, wounds, and infectious diseases. AAF15506 to AAF15514 to
CC AAB57303 represent sequences used in the exemplification of the present
CC invention.
XX
SQ Sequence 2900 BP; 862 A; 635 C; 670 G; 728 T; 5 other;
Query Match 74.1%; Score 2857.4; DB 21; Length 2900;
Best Local Similarity 99.6%; Pred. NO. 0;
Matches 2865; Conservative 4; Mismatches 6; Indels 1; Gaps 1;
QY 25 CAGCGGCTGCAGCAGATTCAGTCCAGCCGCTGCTGAGTACGAGCAGCAGCAGTGTCTG 84
Db 15 CGCGCGGTGCAGAGATTCAGTCCAGCCGCTGCTGAGTACGAGCAGCAGCAGTGTCTG 73
QY -85 GAATGCTCCAGCAGTACGAGGCTAGTATGTCGCGCGCGGCTGCGCGGACCGCTC 144
Db 74 GAATGCTCCAGCAGTACGAGGCTAGTATGTCGCGCGCGGCTGCGCGGACCGCTC 133
QY 145 CTACACACCTTCACAGCTCAGTCCAGCCAGCAGTCCGCTGCTGAGTACGAGCAGCAG 204
Db 134 CTACACACCTTCACAGCTCAGTCCAGCCAGCAGTCCGCTGCTGAGTACGAGCAGCAG 193
QY 205 CAGCGGCGGAGAGAGATTTTATCATCATCAGCTGAAGATAGAGAGAGTGAACACCTC 264
Db 194 CAGCGGCGGAGAGAGATTTTATCATCATCAGCTGAAGATAGAGAGAGTGAACACCTC 253
QY 265 CCTGCGCGGTACAAATGAATCACAAGCAGTGCATAGAGTTCACACAAAGT 324
Db 254 CCTGCGCGGTACAAATGAATCACAAGCAGTGCATAGAGTTCACACAAAGT 313
QY 325 GGTGTATATTTGAGCAAGTAGATCTTGTGCTGACCTCTGACGATAGAAATCCT 384
Db 314 GGTGTATATTTGAGCAAGTAGATCTTGTGCTGACCTCTGACGATAGAAATCCT 373
QY 385 TCAGATTTAATTAAGTGCATCTTGTATAGAGCCACAGAAATATGATCTTGTCAA 444
Db 374 TCAGATTTAATTAAGTGCATCTTGTATAGAGCCACAGAAATATGATCTTGTCAA 433
QY 445 GAAGCATTCATCTTGGCGCTCTTTCGCCAGAAAAAAGAGTGTATTAAGCTTTC 504
Db 434 GAAGCATTCATCTTGGCGCTCTTTCGCCAGAAAAAAGAGTGTATTAAGCTTTC 493
QY 505 ACAGACATGCTGCTTGGCTTGTATGATGTTTGTATGAGAAAGATAGAAAT 564
Db 494 ACAGACATGCTGCTTGGCTTGTATGATGTTTGTATGAGAAAGATAGAAAT 553
QY 565 CTTTGTGAGGAAGATGATCTGAGCAAGTTCATGAGAGATAGAAATCTATTTT 624
Db 554 CTTTGTGAGGAAGATGATCTGAGCAAGTTCATGAGAGATAGAAATCTATTTT 613
QY 625 GAACAGCAAAACCTGAAGTTGTAGAAATCCATGTTCTATGACACTACATGCTTCT 684

Db 614 GAACAGCAAAACCTGAAGTTGTAGAAATCCATGTTCTATGACACCTACATGCTTCT 673
QY 685 ATACAGACTGCTATACAGTACGACATTTTAATGCAATGCTTAAGAGAACTAAATGCGCATAC 744
Db 674 ATACAGACTGCTATACAGTACGACATTTTAATGCAATGCTTAAGAGAACTAAATGCGCATAC 733
QY 745 CCATCGCTTGAAGTGAAGATTTATCTTGAAGAAATGCTATGAGAAACCTTTGAGCAAG 804
Db 734 CCATCGCTTGAAGTGAAGATTTATCTTGAAGAAATGCTATGAGAAACCTTTGAGCAAG 793
QY 805 ACAATCCGCCATTAATCTGATCCTTTGTGCGCAGCTTGGAGCCAAAGACTAAATCCTTA 864
Db 794 ACAATCCGCCATTAATCTGATCCTTTGTGCGCAGCTTGGAGCCAAAGACTAAATCCTTA 853
QY 865 GTTAGAGATTTGAAGATTAATGCAAGCTTGTGAGATCTCTCTGATGATGTTGTC 924
Db 854 GTTAGAGATTTGAAGATTAATGCAAGCTTGTGAGATCTCTCTGATGATGTTGTC 913
QY 925 ACATTTCTTAATCTTCTGGAATCTCTGAGCAAGCAAGCAAGCAAGCTTTGGTCAAGATTC 984
Db 914 ACATTTCTTAATCTTCTGGAATCTCTGAGCAAGCAAGCAAGCAAGCTTTGGTCAAGATTC 973
QY 985 GGTGGCTGTTTCTTGTGACTCCAGCCTCGATGTTTATTAATGCTCGAGCAAGGTTTAT 1044
Db 974 GGTGGCTGTTTCTTGTGACTCCAGCCTCGATGTTTATTAATGCTCGAGCAAGGTTTAT 1033
QY 1045 CATCTCCAGATGCGCAAAATAGTAAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCA 1104
Db 1034 CATCTCCAGATGCGCAAAATAGTAAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCA 1093
QY 1105 GAAGGGGAAGCAAAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCA 1164
Db 1094 GAAGGGGAAGCAAAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCA 1153
QY 1165 ACTGAAGTATTAAGAAATTTGAGGCAAGAAATTAAGAGATGAGTGAAGCTTGTGGTCC 1224
Db 1154 ACTGAAGTATTAAGAAATTTGAGGCAAGAAATTAAGAGATGAGTGAAGCTTGTGGTCC 1213
QY 1225 GGTCAAGTACTGATTTGTGCAAGTATGACCAAGTATGTTCCAGTCCAGAGATATATC 1284
Db 1214 GGTCAAGTACTGATTTGTGCAAGTATGACCAAGTATGTTCCAGTCCAGAGATATATC 1273
QY 1285 ACTCTTGAGCGGAGGCTCTTATTAAGGCTCTACAGGAAACCTTTGAGAGGATATG 1344
Db 1274 ACTCTTGAGCGGAGGCTCTTATTAAGGCTCTACAGGAAACCTTTGAGAGGATATG 1333
QY 1345 AAAGCTGAGAGTCTGATGATTAATTTAGAGAGAGAGAGTTCAAAGAGATTAAGAGAA 1404
Db 1334 AAAGCTGAGAGTCTGATGATTAATTTAGAGAGAGAGAGTTCAAAGAGATTAAGAGAA 1393
QY 1405 TCTCAGAAAAGACCTAAAGACCCCAAAACAAAGAGGCTTCTACCAAGAAAGAAC 1464
Db 1394 TCTCAGAAAAGACCTAAAGACCCCAAAACAAAGAGGCTTCTACCAAGAAAGAAC 1453
QY 1465 CTCAAAAAAGAAAAAGGAAAGTGAACCTTAACCTGAGGAAAGTGAAGAACTG 1524
Db 1454 CTCAAAAAAGAAAAAGGAAAGTGAACCTTAACCTGAGGAAAGTGAAGAAAGTGAAGAACTG 1513
QY 1525 GAAGAGAGAGAGATGTCGAGAGAGAGATATGTCGAGAGAAATTAACAGTACGCCAGAAAGC 1584
Db 1514 GAAGAGAGAGAGATGTCGAGAGAGAGATATGTCGAGAGAAATTAACAGTACGCCAGAAAGC 1573
QY 1585 TGCCCGGAGAAATTAAGCATGAAGATTTGATGATTAATTTGATGAGATGCTGCTTTC 1644
Db 1574 TGCCCGGAGAAATTAAGCATGAAGATTTGATGATTAATTTGATGAGATGCTGCTTTC 1633
QY 1645 GGAATCTGAAAGAACCCCTACTATCATCATCCGCTTCTGCTGAGAGGAGCCCTAT 1704
Db 1634 GGAATCTGAAAGAACCCCTACTATCATCATCCGCTTCTGCTGAGAGGAGCCCTAT 1693
QY 1705 GCTCTGCAAGGATACATGAAGTGAAGGAGCAAGATACGTTGTTTATGAGCAGAG 1764

Db 1694 GCTTGACAAAGGTACTACATGAGTGGAGCCAGATACGGTCTTTATGACGAGAG 1753
Qy 1765 CTAACTTGTTCGGCAGCTTGAATTTACAGGCGAGTAGGCTCGGGAACCTCGTAGG 1824
Db 1754 CTAACTTGTTCGGCAGCTTGAATTTACAGGCGAGTAGGCTCGGGAACCTCGTAGG 1813
Qy 1825 GTTACTTCTTATATACGAGGAGTTCACATGAGGAACAAGCTATCTCAGCTTTGCGG 1884
Db 1814 GTTACTTCTTATATACGAGGAGTTCACATGAGGAACAAGCTATCTCAGCTTTGCGG 1873
Qy 1885 AAAGAAAAGAAAGCTTTGAAAAAATCATAGGAAAAAGCAAGATGGTTCCCTGAA 1944
Db 1874 AAAGAAAAGAAAGCTTTGAAAAAATCATAGGAAAAAGCAAGATGGTTCCCTGAA 1933
Qy 1945 GAAAGAGAAGGAGAGATGAACAACCTAGACCTAGTAAGAGGACACAGCATCTGCAGT 2004
Db 1934 GAAAGAGAAGGAGAGATGAACAACCTAGACCTAGTAAGAGGACACAGCATCTGCAGT 1993
Qy 2005 GTTTCACACTGACACTCGGAAAGCCGGTGGCCAGAAACAGATGTAACAGCAAGACATA 2064
Db 1994 GTTTCACACTGACACTCGGAAAGCCGGTGGCCAGAAACAGATGTAACAGCAAGACATA 2053
Qy 2065 GTTGTGATATCGTGAAATTTGAAAGTGAGCTTCCATCTGTATCCATCGTGGGGCAT 2124
Db 2054 GTTGTGATATCGTGAAATTTGAAAGTGAGCTTCCATCTGTATCCATCGTGGGGCAT 2113
Qy 2125 GACATTGAACCCGCTTAAAGTGGAGATTAATCCTCTACCTCCAGAAATGTCGGT 2184
Db 2114 GACATTGAACCCGCTTAAAGTGGAGATTAATCCTCTACCTCCAGAAATGTCGGT 2173
Qy 2185 GAGCCCAAGAGATATCAGTATTAATCGCTTTAAATAAAGCGCCGCTACAGCCAG 2244
Db 2174 GAGCCCAAGAGATATCAGTATTAATCGCTTTAAATAAAGCGCCGCTACAGCCAG 2233
Qy 2245 TGATCTCATATCCCGCTACTACAAAGCGTCCCGTCTGTATGAGTTGACCTTAGC 2304
Db 2234 TGATCTCATATCCCGCTACTACAAAGCGTCCCGTCTGTATGAGTTGACCTTAGC 2293
Qy 2305 AAGCTTTCTCTCTACTTCCCGAGTGGCTGTTCCAGGAGATCCAGCAATGACAT 2364
Db 2294 AAGCTTTCTCTCTACTTCCCGAGTGGCTGTTCCAGGAGATCCAGCAATGACAT 2353
Qy 2365 AGTTCCAACTCATCTTCTTAAACACTTCCCGAGACTAGGATCTCTGTGCCCC 2424
Db 2354 AGTTCCAACTCATCTTCTTAAACACTTCCCGAGACTAGGATCTCTGTGCCCC 2413
Qy 2425 TCTCTCATGCAAGCGCGGAGTTGTTGAGAGCTGAAACAAGCAAGCCACAGCTGAT 2484
Db 2414 TCTCTCATGCAAGCGCGGAGTTGTTGAGAGCTGAAACAAGCAAGCCACAGCTGAT 2473
Qy 2485 GCGGCGAAGAGCATGGCCATTACAGCAGATTCGAAACCCCTCCGAGTCAGAGAGAT 2544
Db 2474 GCGGCGAAGAGCATGGCCATTACAGCAGATTCGAAACCCCTCCGAGTCAGAGAGAT 2533
Qy 2545 AATCTGTGTCGCCAAGACTTCTTTTAAAAATGCCAGGGGTGAATGCCAAAAATGCGCGC 2604
Db 2534 AATCTGTGTCGCCAAGACTTCTTTTAAAAATGCCAGGGGTGAATGCCAAAAATGCGCGC 2593
Qy 2605 TCCTTGATGACACACGCTTAAGAACATCGCAGATTAGACGCCCTTCACAGACAGACTC 2664
Db 2594 TCCTTGATGACACACGCTTAAGAACATCGCAGATTAGACGCCCTTCACAGACAGACTC 2653
Qy 2665 ACGATATTTGSGGAGATGCTGCAAAATGCCAAACAGCTTTATGATTTATTCACACCTCT 2724
Db 2654 ACGATATTTGSGGAGATGCTGCAAAATGCCAAACAGCTTTATGATTTATTCACACCTCT 2713
Qy 2725 TTTGCGAAGTCTATCAAAAGGAAAAAGTGAACAGTATGCTGTTTCTTA 2784
Db 2714 TTTGCGAAGTCTATCAAAAGGAAAAAGTGAACAGTATGCTGTTTCTTA 2773
Qy 2785 TCCCATGCTGTACTTTTCAGCGGCTCTTGCCAGACATCATAGTCAATTATTAATTAT 2844
Db 2774 TCCCATGCTGTACTTTTCAGCGGCTCTTGCCAGACATCATAGTCAATTATTAATTAT 2833

Qy 2845 GGTTCGATATTTCAATCTTTCCAAATGCTTAATGATTGTACGGTGGACAGAGCCAG 2904
Db 2834 GGTTCGATATTTCAATCTTTCCAAATGCTTAATGATTGTACGGTGGACAGAGTTGAG 2893

RESULT 4
AAS63230
ID AAS63230 standard; cDNA; 2718 BP.

XX AAS63230;

DT 29-JAN-2002 (first entry)

XX Human full-length XPF (ERCC4) DNA.

XX DNA mutation-binding protein; nuclease; DNA mismatch; cancer; PCR primer;
XX DNA damage; human xeroderma pigmentosum complementation group; XPF; XPA;
XX XPC; XPE; ERCC4; human Muts homologue 2; hMSH2; Muts; Nuc; MutY; Fpg; ss;
XX Fapy-DNA glycosylase; uracil DNA glycosylase; ung; Tdg; xthA gene; Uvr A;
XX A/G-specific adenine glycosylase; synthetic T4 endonuclease V; T4 endo V;
XX thymine DNA-glycosylase; Uvr B; Uvr C; nth gene; nfo gene; exonuclease;
XX endonuclease.

OS Homo sapiens.

XX WO200173079-A2.

XX 04-OCT-2001.

XX 26-MAR-2001; 2001WO-US09700.

XX 28-MAR-2000; 2000US-192764P.

XX 29-AUG-2000; 2000US-0650855.

XX (REGC) UNIV CALIFORNIA.

XX Mc Cutchen-maloney SL;

XX WPI; 2001-656920/75.

XX P-PSDB; AA069743.

XX Recombinant chimeric protein, useful for detecting and quantifying DNA
XX mutations, e.g. in disease diagnosis, comprises mutation-binding
XX protein and nuclease

XX Disclosure; Page 69-70; 128p; English.

XX Sequences AAS63224-AAS63305 represent cDNA encoding and PCR primers used
XX to amplify cDNA encoding proteins which can be used in the synthesis of
XX chimeric proteins comprising a DNA mutation-binding protein, a linker and
XX a nuclease, by recombinant technology. The chimeric proteins are useful
XX for detection, quantification and mapping of DNA sequence variations
XX including mutations, for example, caused by damage and mismatches. The
XX proteins are able to bind to the site of the DNA mutation and cut it out
XX of the molecule. This is useful for early diagnosis of cancer and other
XX diseases. The proteins used in the invention include human XPF (or
XX ERCC4), human xeroderma pigmentosum complementation groups A, C and E
XX (XPA, XPC and XPE), human Muts homologue 2 (hMSH2), *Serratia marcescens*
XX nuclease (Nuc), *Thermus thermophilus* Muts, *Escherichia coli* Fapy-DNA
XX glycosylase (Fpg), uracil DNA glycosylase (ung), A/G-specific adenine
XX glycosylase (MutY), synthetic T4 endonuclease V (T4 endo V), thymine
XX DNA-glycosylase (Tdg), *E. coli* Uvr A, B and C, and *E. coli* endonucleases
XX and exonucleases.

XX Sequence 2718 BP; 827 A; 588 C; 634 G; 669 T; 0 other;

XX Query Match 70.5%; Score 2718; DB 22; Length 2718;

XX Best Local Similarity 100.0%; Pred. No. 0; Mismatches 0; Gaps 0;

XX Matches 2718; Conservative 0; Indels 0; Gaps 0;

Qy 46 ATGGCGCGCTGCTGAGTACAGCAGCAGCTGTGTGGAAGTCTCTGACACTGACGG 105
|||||

Db 1 ATGGCCCGCTGCTGGAGTACGACGACAGCTGTGCTGGAACTGCTGACACTGACGCG 60
QY 106 CTAGTAGTGTGCG 165
Db 61 CTAGTAGTGTGCG 120
QY 166 CACTGGCACCACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 225
Db 121 CACTGGCACCACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 180
QY 226 TTTATCATCATGCTGAAGATAGAGAGAGTGAACACCTCCCTCGCGCTGTAAACAAATGAA 285
Db 181 TTTATCATCATGCTGAAGATAGAGAGAGTGAACACCTCCCTCGCGCTGTAAACAAATGAA 240
QY 286 ATACAAAGCAACAGTGCCTGTGAAGTTTACACAAAGTGTGTATATTTGGCAGCAAGT 345
Db 241 ATACAAAGCAACAGTGCCTGTGAAGTTTACACAAAGTGTGTATATTTGGCAGCAAGT 300
QY 346 AGGATACCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 405
Db 301 AGGATACCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 360
QY 406 TTGGTGTATGAGCCGACCAATATAGAGTCTTGTCAAGACATTCATCTGCGCTC 465
Db 361 TTGGTGTATGAGCCGACCAATATAGAGTCTTGTCAAGACATTCATCTGCGCTC 420
QY 466 TTTCGGCAGAAAAACAAGCTGTTTTATTAAAGCTTTCACAGACATTCGCTGCTT 525
Db 421 TTTCGGCAGAAAAACAAGCTGTTTTATTAAAGCTTTCACAGACATTCGCTGCTT 480
QY 526 GATACGTGTTTTGTCACTGTGAGAAAGATGAGAAATCTTTTTGTGAGAAACCTGTAT 585
Db 481 GATACGTGTTTTGTCACTGTGAGAAAGATGAGAAATCTTTTTGTGAGAAACCTGTAT 540
QY 586 CTGTGGCCAAAGTTCATGTAGCAGTAACTCATTTTAAACAGCAACCACTGAAAGTT 645
Db 541 CTGTGGCCAAAGTTCATGTAGCAGTAACTCATTTTAAACAGCAACCACTGAAAGTT 600
QY 646 GTAGAAATCCATGTTTCTATGACACCTACCATGCTGCTATACAGATTCGCTATCTGGAC 705
Db 601 GTAGAAATCCATGTTTCTATGACACCTACCATGCTGCTATACAGATTCGCTATCTGGAC 660
QY 706 ATTTTAAATCATGTCTAAAGAACTAAATGCCATTAACCATTCGCTTGAAGTGAAGAT 765
Db 661 ATTTTAAATCATGTCTAAAGAACTAAATGCCATTAACCATTCGCTTGAAGTGAAGAT 720
QY 766 TTTATCTTAAATAATGCTATATGAAAACTTTTGACAAAGCAATCCGCCATATCTGAT 825
Db 721 TTTATCTTAAATAATGCTATATGAAAACTTTTGACAAAGCAATCCGCCATATCTGAT 780
QY 826 CCTTGTGGCAGCAGCTTGGAGCCAACTAAATCCTTAGTTCAGATTTGAAGATATTA 885
Db 781 CCTTGTGGCAGCAGCTTGGAGCCAACTAAATCCTTAGTTCAGATTTGAAGATATTA 840
QY 886 CGAATTTGCTGCAGTATCTCTCACTATGATGTGTCAATTTCTTAACTCTTCTGAAA 945
Db 841 CGAATTTGCTGCAGTATCTCTCACTATGATGTGTCAATTTCTTAACTCTTCTGAAA 900
QY 946 TCTCTGAGACAAAGGAAAAAGCTTTTGTGACAAATCAAGTTGGCTGTTCTTGACTCC 1005
Db 901 TCTCTGAGACAAAGGAAAAAGCTTTTGTGACAAATCAAGTTGGCTGTTCTTGACTCC 960
QY 1006 AGACACCTGATGTTTAAATATGCTCGAGCAAGGTTTATCATCTTCCAGATGCCAAATG 1065
Db 961 AGACACCTGATGTTTAAATATGCTCGAGCAAGGTTTATCATCTTCCAGATGCCAAATG 1020
QY 1066 AGTAAAAAAGAAAAATATCTGAAAAATGAAATTAAGAAAGGGAAGAAACAAAAAG 1125
Db 1021 AGTAAAAAAGAAAAATATCTGAAAAATGAAATTAAGAAAGGGAAGAAACAAAAAG 1080
QY 1126 GAAGTGTCTAGAAAGCAACCAAGTGGAGGCACTGACTGAGATTTAAAGAAATTT 1185
Db 1081 GAAGTGTCTAGAAAGCAACCAAGTGGAGGCACTGACTGAGATTTAAAGAAATTT 1140

QY 1186 GAGCAGAAAAATAAGAGAGGAAAGCTTGTGTGCTCAGAGTCAACTACATGATTTGTGCA 1245
Db 1141 GAGCAGAAAAATAAGAGAGTGAAGACTTGTGTGCTCAGAGTCAACTACATGATTTGTGCA 1200
QY 1246 AGTATGACCGAACAATGTTCCAGCTGAGAGTATATCACTCTTGTGAGCGGAGCCCTTC 1305
Db 1201 AGTATGACCGAACAATGTTCCAGCTGAGAGTATATCACTCTTGTGAGCGGAGCCCTTC 1260
QY 1306 TTTATGAGGCTCTACAGAAAAACCTTTGAGAAGATAGCAAAAGCTGAAGAAAGTGTGATG 1365
Db 1261 TTTATGAGGCTCTACAGAAAAACCTTTGAGAAGATAGCAAAAGCTGAAGAAAGTGTGATG 1320
QY 1366 AAATTTGAGAAAGAAAGCACTTCAAAAGAAATTTAGAAATCTCAAAAAGACCTTAAAGAC 1425
Db 1321 AAATTTGAGAAAGAAAGCACTTCAAAAGAAATTTAGAAATCTCAAAAAGACCTTAAAGAC 1380
QY 1426 CCCCCAAAACAAGACGGGCTTCTACCAAAAGAAACCTCAAAAAGAAACGGAAG 1485
Db 1381 CCCCCAAAACAAGACGGGCTTCTACCAAAAGAAACCTCAAAAAGAAACGGAAG 1440
QY 1486 TTGACCTTAACCTCAATGATAGGAAACCTGAAGAACTGGAAGGGAAGAGATGTGAG 1545
Db 1441 TTGACCTTAACCTCAATGATAGGAAACCTGAAGAACTGGAAGGGAAGAGATGTGAG 1500
QY 1546 GAAAGATATGCTGAGAAATTAAGCAGTAGCCAGAAAGCTGCCGGAAGAAATTAACAT 1605
Db 1501 GAAAGATATGCTGAGAAATTAAGCAGTAGCCAGAAAGCTGCCGGAAGAAATTAACAT 1560
QY 1606 GAAAGATTTGATGTAATTTGTCTATGCGAGTGTGCTTTGGAATCTCTGAAGAAACCTTC 1665
Db 1561 GAAAGATTTGATGTAATTTGTCTATGCGAGTGTGCTTTGGAATCTCTGAAGAAACCTTC 1620
QY 1666 ACTATCATCATCGCTGCTGCGGTTGAGAGGACCCCTATGCTGACAAAGGATCTCAT 1725
Db 1621 ACTATCATCATCGCTGCTGCGGTTGAGAGGACCCCTATGCTGACAAAGGATCTCAT 1680
QY 1726 GAAAGTGAAGCAGAGATACGTGTTCTTTATGACGACAGAGTAACTTTGTTTGGCAGCTT 1785
Db 1681 GAAAGTGAAGCAGAGATACGTGTTCTTTATGACGACAGAGTAACTTTGTTTGGCAGCTT 1740
QY 1786 GAAATTTTACAGGCGAGTAGGCGCTGGGAAACCTCTGAGGCTTTTACTTTCTTATATACGA 1845
Db 1741 GAAATTTTACAGGCGAGTAGGCGCTGGGAAACCTCTGAGGCTTTTACTTTCTTATATACGA 1800
QY 1846 GGTTCACATGAGGAAACAGCTATCTCATGCTTTGSGGAAAGAAAGAAAGCTTTGAA 1905
Db 1801 GGTTCACATGAGGAAACAGCTATCTCATGCTTTGSGGAAAGAAAGAAAGCTTTGAA 1860
QY 1906 AAATCTATTAAGGAAAAAGCAAGCATGTTGCTCCCTGAGAAAGAAAGCAAGCAGAGATGAA 1965
Db 1861 AAATCTATTAAGGAAAAAGCAAGCATGTTGCTCCCTGAGAAAGAAAGCAAGCAGAGATGAA 1920
QY 1966 ACAAACTTGAACCTAGTAAAGGACACAGCATCTGACAGATGTTTCTCACTGACACTCGGAAA 2025
Db 1921 ACAAACTTGAACCTAGTAAAGGACACAGCATCTGACAGATGTTTCTCACTGACACTCGGAAA 1980
QY 2026 GCCGCTGGCCAGGAAGCAAGATGATACAGCAAGCAATGATGTGGAATGATGCGGAATTT 2085
Db 1981 GCCGCTGGCCAGGAAGCAAGATGATACAGCAAGCAATGATGTGGAATGATGCGGAATTT 2040
QY 2086 CGAAGTGAAGCTTCATCTGATATCATGCTGCGGGCATTTGACATTTGAACCGGTGACTTTA 2145
Db 2041 CGAAGTGAAGCTTCATCTGATATCATGCTGCGGGCATTTGACATTTGAACCGGTGACTTTA 2100
QY 2146 GAGGTGAGATTAATCATCTCACTCCAGAAATGTGCTGAGCGCAAGAGATATCAGTAT 2205
Db 2101 GAGGTGAGATTAATCATCTCACTCCAGAAATGTGCTGAGCGCAAGAGATATCAGTAT 2160
QY 2206 TTTATGCGCTCTTAAATTAACGGCGGCTCTACAGCCAGTGCATCTCATGTCGCGCTAC 2265
Db 2161 TTTATGCGCTCTTAAATTAACGGCGGCTCTACAGCCAGTGCATCTCATGTCGCGCTAC 2220

Db 30242 GAAGTCGATCAAAAGGAAAGGAAAGATGACAGTGGCTGTTTCTATCCAT 30301
OY 2791 GCCTGACTTTTCACGGGCTCCTTGCAGACATCATAGTCTATTATATTGTTG 2850
Db 30302 GCCTGACTTTTCACGGGCTCCTTGCAGACATCATAGTCTATTATATTGTTG 30361
OY 2851 CTAATTCATCTCTTTCCAAATGCTCTTAATGATGTACGGTGACCAAGACGAGATTCC 2910
Db 30362 CTAATTCATCTCTTTCCAAATGCTCTTAATGATGTACGGTGACCAAGACGAGATTCC 30421
OY 2911 TCTCTGAACCTGACAGTTAGGACATCATTTGACTGCTGCTCTTTTCTCTCC 2970
Db 30422 TCTCTGAACCTGACAGTTAGGACATCATTTGACTGCTGCTCTTTTCTCTCC 30481
OY 2971 TGCACCGCTATAGCCGGGCTAGCATGTTCTTTTAATGAGGTTTGACAGATCAGGT 3030
Db 30482 TGCACCGCTATAGCCGGGCTAGCATGTTCTTTTAATGAGGTTTGACAGATCAGGT 30541
OY 3031 AAAGTTCTACAGATGATACAGAAAGTAAAGCTTTTACCTATCTTACAGATCTCAT 3090
Db 30542 AAAGTTCTACAGATGATACAGAAAGTAAAGCTTTTACCTATCTTACAGATCTCAT 30601
OY 3091 TAGAAAGAAATATGCTAAGCTTGATGAGCGGTGACGGGAGGAAAGACAGCGACAA 3150
Db 30602 TAGAAAGAAATATGCTAAGCTTGATGAGCGGTGACGGGAGGAAAGACAGCGACAA 30661
OY 3151 GAAAGTACCATTTTAAACAGTCTTGTATCTAGTGAACAATTAATACATCTTAAT 3210
Db 30662 GAAAGTACCATTTTAAACAGTCTTGTATCTAGTGAACAATTAATACATCTTAAT 30721
OY 3211 GCACCTATACCATGCTGCTGCTGCTCCAAATCTGGTCTTGTCTGCTGCTGA 3270
Db 30722 GCACCTATACCATGCTGCTGCTGCTCCAAATCTGGTCTTGTCTGCTGCTGA 30781
OY 3271 CGCTTGAACTGATGTTTGTGTAGGAATCATGTTCTGACCCCTTGTCTAACAAGACCT 3330
Db 30782 CGCTTGAACTGATGTTTGTGTAGGAATCATGTTCTGACCCCTTGTCTAACAAGACCT 30841
OY 3331 TCTGGAACCTGAGAAAGAAACATCTTGGCATTCCTGACAGTCTCTACACAT 3390
Db 30842 TCTGGAACCTGAGAAAGAAACATCTTGGCATTCCTGACAGTCTCTACACAT 30901
OY 3391 TTTCTGAGCTCATACTTCTGCTGCTGCTCTAAGAAATTTTCAGAGCCCTCTACT 3450
Db 30902 TTTCTGAGCTCATACTTCTGCTGCTGCTCTAAGAAATTTTCAGAGCCCTCTACT 30961
OY 3451 ACTAATTCAGAGAGTCTCTCTCAAAACGTGGTGAATCTTCTTAATGACCTTAACATA 3510
Db 30962 ACTAATTCAGAGAGTCTCTCTCAAAACGTGGTGAATCTTCTTAATGACCTTAACATA 31021
OY 3511 TGTAGCATTAATTAATTTTCAATTTGTTCCAAATTAATTTTAAAGCAAAATGAATTAC 3570
Db 31022 TGTAGCATTAATTAATTTTCAATTTGTTCCAAATTAATTTTAAAGCAAAATGAATTAC 31081
OY 3571 CTGTTTGCAAAAGTAAATGATGAGAGGCTCTTGAATTTCTCAATTTTTCACATATCA 3630
Db 31082 CTGTTTGCAAAAGTAAATGATGAGAGGCTCTTGAATTTCTCAATTTTTCACATATCA 31141
OY 3631 GTTCTCAATATACAGATTCCTCAATCTCAATCTGAGTGTAAAGAGTTTTCACACT 3690
Db 31142 GTTCTCAATATACAGATTCCTCAATCTCAATCTGAGTGTAAAGAGTTTTCACACT 31201
OY 3691 CCGTGTCTCAGCTCTTATATCTTAAGACACAGCATCATATCTTGAATTAACAC 3750
Db 31202 CCGTGTCTCAGCTCTTATATCTTAAGACACAGCATCATATCTTGAATTAACAC 31261
OY 3751 TAAATGGCAGTGAAGCGAGATGCGACACTGCAACCCCTGCTGCGGAGACAGACTGAGCT 3810
Db 31262 TAAATGGCAGTGAAGCGAGATGCGACACTGCAACCCCTGCTGCGGAGAGAGTGAAGCT 31321
OY 3811 TTGTCTCTATTAACAAAAGAAAAGAAAAGAAATACAACTTAACCTCA 3857
Db 31322 TTGTCTCTATTAACAAAAGAAAAGAAAAGAAATACAACTTAACCTCA 31368

RESULT 6
ID ABL64428 standard: DNA: 35641 BP.
AC ABL64428;
DT 15-MAY-2002 (first entry)
XX Stomach cancer related gene sequence SEQ ID NO:2765.
DE Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;
XX stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;
KW cytostatic; gene therapy; antineoplastic; Wilms tumour; adenocarcinoma;
KM gene; ds.
XX
OS Homo sapiens.
XX
PN W0200194629-A2.
XX
PD 13-DEC-2001.
XX
PF 30-MAY-2001; 2001WO-US10838.
XX
PR 05-JUN-2000; 2000US-209473P.
PR 18-SEP-2000; 2000US-209531P.
PR 18-SEP-2000; 2000US-233133P.
PR 18-SEP-2000; 2000US-233617P.
PR 20-SEP-2000; 2000US-234009P.
PR 20-SEP-2000; 2000US-234034P.
PR 20-SEP-2000; 2000US-234052P.
PR 22-SEP-2000; 2000US-234509P.
PR 22-SEP-2000; 2000US-234567P.
PR 25-SEP-2000; 2000US-234923P.
PR 25-SEP-2000; 2000US-234924P.
PR 25-SEP-2000; 2000US-235077P.
PR 25-SEP-2000; 2000US-235082P.
PR 25-SEP-2000; 2000US-235134P.
PR 25-SEP-2000; 2000US-235280P.
PR 26-SEP-2000; 2000US-235637P.
PR 26-SEP-2000; 2000US-235638P.
PR 27-SEP-2000; 2000US-235711P.
PR 27-SEP-2000; 2000US-235720P.
PR 27-SEP-2000; 2000US-235840P.
PR 27-SEP-2000; 2000US-235863P.
PR 28-SEP-2000; 2000US-236028P.
PR 28-SEP-2000; 2000US-236032P.
PR 28-SEP-2000; 2000US-236033P.
PR 28-SEP-2000; 2000US-236034P.
PR 28-SEP-2000; 2000US-236109P.
PR 28-SEP-2000; 2000US-236111P.
PR 29-SEP-2000; 2000US-236842P.
PR 29-SEP-2000; 2000US-236891P.
PR 02-OCT-2000; 2000US-237172P.
PR 02-OCT-2000; 2000US-237173P.
PR 02-OCT-2000; 2000US-237278P.
PR 02-OCT-2000; 2000US-237294P.
PR 02-OCT-2000; 2000US-237295P.
PR 02-OCT-2000; 2000US-237316P.
PR 03-OCT-2000; 2000US-237425P.
PR 03-OCT-2000; 2000US-237598P.
PR 03-OCT-2000; 2000US-237604P.
PR 03-OCT-2000; 2000US-237606P.
PR 03-OCT-2000; 2000US-237608P.
PR 01-NOV-2000; 2000US-244667P.
PR 01-NOV-2000; 2000US-245084P.
XX
XX (AVAL-) AVALON PHARM.
XX
PI Young PE, Augustus M, Carter KC, Edner R, Endress G, Horrigan S;
XX Soppet DR, Weaver Z;

DR WPI: 2002-188264/24.
XX Screening for anti-neoplastic agent involves exposing cells to a
PT chemical agent to be tested for anti-neoplastic activity, and
XX determining a change in expression of a gene of a signature gene set
PS Claim 1: SEQ ID 2765; 44p; English.
XX
CC The present invention describes a method (M1) for screening for an
CC anti-neoplastic agent. The method involves exposing cells to a chemical
CC agent to be tested for anti-neoplastic activity, determining a change in
CC expression of at least one gene (I) of a signature gene set, where (I)
CC comprises a sequence (S) selected from 8447 sequences (given in AB161664
CC to AB170110), or is at least 95% identical to (S), where a change in
CC expression is indicative of anti-neoplastic activity. (I) has cytostatic
CC activity and can be used in gene therapy. M1 can be used for screening
CC an anti-neoplastic agent, and can be used for producing a product which
CC is the data collected with respect to the anti-neoplastic agent as a
CC result of M1, and the data is sufficient to convey the chemical
CC structure and/or properties of the agent. M1 can be used in the
CC treatment of cancer such as colon, breast, stomach, lung, thyroid,
CC osteophagel, ovarian, kidney, prostate or pancreatic cancer,
CC adenocarcinoma, carcinoma, clear cell cancer, infiltrating ductal cancer,
CC infiltrating lobular cancer, squamous cell carcinoma, neuroendocrine
CC carcinoma, papillary carcinoma and Wilms' tumour.
XX
SQ Sequence 35641 BP; 10575 A; 7169 C; 7323 G; 10574 T; 0 other;
Query Match 47.4%; Score 1829.4; DB 24; Length 35641;
Best Local Similarity 99.4%; Pred. No. 0;
Matches 183; Conservative 0; Mismatches 11; Indels 0; Gaps 0;
QY 2011 ACTGACACTCGGAAGCGGTGGCCAGAGAGATGTATACAGCAAGCATATGTTG 2070
DB 29522 ACTTACTTTTCTCTGTAGTGGCCAGAGACAGATGTTACACAGCAAGCATATGTTG 29521
QY 2071 GATATGCGTGAATTTGGAAGTGAAGTTCATCTCTGATCCATCGTGGGGATTGACAT 2130
DB 29582 GATATGCGTGAATTTGGAAGTGAAGTTCATCTCTGATCCATCGTGGGGATTGACAT 29641
QY 2131 GAACCGGAGCTTGAAGTGGAGATTACATCTCTGATCCATCGAAGAAATTTGGTGGAGCG 2190
DB 29642 GAACCGGAGCTTGAAGTGGAGATTACATCTCTGATCCATCGAAGAAATTTGGTGGAGCG 29701
QY 2191 AAGATATCATGATTTAATTAATGAGTTCCTTAATTAAGGCGCCCTCTACAGCAGATC 2250
DB 29702 AAGATATCATGATTTAATTAATGAGTTCCTTAATTAAGGCGCCCTCTACAGCAGATC 29761
QY 2251 TGCATGTCGCCCTACTACAAAGCGTCCGCTGCTGATTTGATTTGACCTTAGCAAGCT 2310
DB 29762 TGCATGTCGCCCTACTACAAAGCGTCCGCTGCTGATTTGATTTGACCTTAGCAAGCT 29821
QY 2311 TTCTCTCTCACTTCCCGAGTGCCTTGTTCAGAGATCTCCAGCAATGACATTAATTC 2370
DB 29822 TTCTCTCTCACTTCCCGAGTGCCTTGTTCAGAGATCTCCAGCAATGACATTAATTC 29881
QY 2371 AAATCATCTCTTTTACACTCTACCTCCAGACTAGAGATTTCTGTGGTCCCTCTCT 2430
DB 29882 AAATCATCTCTTTTACACTCTACCTCCAGACTAGAGATTTCTGTGGTCCCTCTCT 29941
QY 2431 CATGCAACGGGGAGTGTGTTGAGAGCTGAAGCAAGCAAGCAAGCAAGCTGATGCGCG 2490
DB 29942 CATGCAACGGGGAGTGTGTTGAGAGCTGAAGCAAGCAAGCAAGCAAGCTGATGCGCG 30001
QY 2491 ACAGCACTGGCATTAACAGCAGATTTGAAACCTTCCCGAGTCAAGAGAAATTAATCT 2550
DB 30002 ACAGCACTGGCATTAACAGCAGATTTGAAACCTTCCCGAGTCAAGAGAAATTAATCT 30061
QY 2551 GGTCGCCAAGACTCTTGTAAAAAATGCGAGGGTGAATGCAAAATCTGCCCTCTTG 2610
DB 30062 GGTCGCCAAGACTCTTGTAAAAAATGCGAGGGTGAATGCAAAATCTGCCCTCTTG 30121
QY 2611 ATGCACCAAGTTAAGAACATGCGAGAAATTAGCAGCCTGTTCACAAAGAGACCTCAGAGT 2670

DB 30122 ATGCACCAAGTTAAGAACATGCGAGAAATTAGCAGCCTGTTCACAAAGAGACCTCAGAGT 30181
QY 2671 ATTCGGGGAATGCTGCAAAATGCCAAACAGCTTTATGATTTTCATTCACACCTCTTTGCA 2730
DB 30182 ATTCGGGGAATGCTGCAAAATGCCAAACAGCTTTATGATTTTCATTCACACCTCTTTGCA 30241
QY 2731 GAAGTCGATCAAAAGGAAAGGAAAGGAAAGTGAACAGATGCGCTTTCTATCCAT 2790
DB 30242 GAAGTCGATCAAAAGGAAAGGAAAGGAAAGTGAACAGATGCGCTTTCTATCCAT 30301
QY 2791 GCGTCTACTTTTCAAGCGGCTCTCTCCAGACATCATATGATTAATTAATTTGTTG 2850
DB 30302 GCGTCTACTTTTCAAGCGGCTCTCTCCAGACATCATATGATTAATTAATTTGTTG 30361
QY 2851 CTATTTCAATCTTTTCCAAATGCTCTTAATGATTTGACGGTGAACCAAGCCAGATTC 2910
DB 30362 CTATTTCAATCTTTTCCAAATGCTCTTAATGATTTGACGGTGAACCAAGCCAGATTC 30421
QY 2911 TCTGTGAACCTGCGAGTTAAGGATGATCACTTGAACCTGCGCTCTTTCTCTCC 2970
DB 30422 TCTGTGAACCTGCGAGTTAAGGATGATCACTTGAACCTGCGCTCTTTCTCTCC 30481
QY 2971 TGCACCGTCTATGCCGGCTTAGCATGTTCTTTTAAATGAGTTGTCAAGATCAGT 3030
DB 30482 TGCACCGTCTATGCCGGCTTAGCATGTTCTTTTAAATGAGTTGTCAAGATCAGT 30541
QY 3031 AAAGTCTCTACAGATGATTTACAGAGATGAGAACTTACCTGATCTTACAGATCTCAT 3090
DB 30542 AAAGTCTCTACAGATGATTTACAGAGATGAGAACTTACCTGATCTTACAGATCTCAT 30601
QY 3091 TAGAAGGAATATGCTTAAGCCTGCGATGAGCGTCCAGAGGAGGAAAGAGCAGGACAA 3150
DB 30602 TAGAAGGAATATGCTTAAGCCTGCGATGAGCGTCCAGAGGAGGAAAGAGCAGGACAA 30661
QY 3151 GAAAGCTACATTTTAAACAGTCTTGTATCTAGTGAACATAAATTAACAGTCTTAAT 3210
DB 30662 GAAAGCTACATTTTAAACAGTCTTGTATCTAGTGAACATAAATTAACAGTCTTAAT 30721
QY 3211 GCACCTTAACCATGCTCTGTCGCTGCTCCAAATCTGCTGCTTGTGCTGCTCTGGA 3270
DB 30722 GCACCTTAACCATGCTCTGTCGCTGCTCCAAATCTGCTGCTTGTGCTGCTCTGGA 30781
QY 3271 CGCTTGAAGTGTGTTGTGTAAGAAATCATGTTGACCCCTTGTCTACAAAGAGGCT 3330
DB 30782 CGCTTGAAGTGTGTTGTGTAAGAAATCATGTTGACCCCTTGTCTACAAAGAGGCT 30841
QY 3331 TCTGGAACACTGAGAAAGAAACATCTCTTGGCATCTGAGCAGTTCCTCTACACAT 3390
DB 30842 TCTGGAACACTGAGAAAGAAACATCTCTTGGCATCTGAGCAGTTCCTCTACACAT 30901
QY 3391 TTCTTCACTCATATCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3450
DB 30902 TTCTTCACTCATATCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 30961
QY 3451 ACTAATTAAGAGAGTCTCTCAAAAGTGGTGCATGCTCTTAAGACCCCTAACATA 3510
DB 30962 ACTAATTAAGAGAGTCTCTCAAAAGTGGTGCATGCTCTTAAGACCCCTAACATA 31021
QY 3511 TGTAGCATATATCTAATTTATTTATTTGTTCCAAATTTGATTTTAAAGCAAAATGAATAC 3570
DB 31022 TGTAGCATATATCTAATTTATTTGTTCCAAATTTGATTTTAAAGCAAAATGAATAC 31081
QY 3571 CTGTTTGCAGAAAGTTAATGATGAGAGAGCTCTTGAATTTCTCAATTTTGCACATATCA 3630
DB 31082 CTGTTTGCAGAAAGTTAATGATGAGAGAGCTCTTGAATTTCTCAATTTTGCACATATCA 31141
QY 3631 GTCCTCTAATATCAAGAGATTCCTTAAGTCAAGTCCGCTGATGACAGATTTTTCAGACT 3690
DB 31142 GTCCTCTAATATCAAGAGATTCCTTAAGTCAAGTCCGCTGATGACAGATTTTTCAGACT 31201
QY 3691 CCTGCTTCTCAGGCTCTAATCTTAAGACACAGCATATCTCTTGAAGAAATACAC 3750

Db 31202 CCTGCTTCACAGCTCTCTATATCTTAAGACACACGATCATATCTCTGAATATACAC 31261
QY 3751 TAATTGGCAGTGAAGCCGAGATGCGACCACTGCACCCCTGCTGGGAGACAGAGTAGACT 3810
Db 31262 TAATTGGCAGTGAAGCCGAGATGCGACCACTGCACCCCTGCTGGGAGACAGAGTAGACT 31321
QY 3811 TTGTCCTATTTACAAAAGAAAAGAAAAGAAAATACAACTTAAGCTCA 3857
Db 31322 TTGTCCTATTTACAAAAGAAAAGAAAAGAAAATACAACTTAAGCTCA 31368

RESULT 7

AA563229
ID AA563229 standard: cDNA: 1101 BP.

XX AA563229;

DT 29-JAN-2002 (first entry)

XX DNA encoding human XPF (ERCC4) active peptide fragment.

XX DNA mutation-binding protein; nuclease: DNA mismatch; cancer; PCR primer;
KW DNA damage; human xeroderma pigmentosum complementation group; XPF; XPA;
KW XPC; XPE; ERCC4; human Muts homolog 2; hMSH2; Muts; Nuc; Muty; Fpg; ss;
KW Fapy-DNA glycosylase; uracil DNA glycosylase; ung; TDG; xtha gene; Uvr A;
KW A/G-specific adenine glycosylase; synthetic T4 endonuclease V; T4 endo V;
KW thymine DNA-glycosylase; Uvr B; Uvr C; nch gene; nfo gene; exonuclease;
endonuclease.

XX Homo sapiens.

XX W0200173079-A2.

PD 04-OCT-2001.

XX 26-MAR-2001; 2001WO-US09700.

XX 28-MAR-2000; 2000US-192764P.

PR 29-AUG-2000; 2000US-0650855.

XX (REGC) UNIV CALIFORNIA.

XX Mc Cutchen-maloney SL;

XX MPI: 2001-656920/75.

DR P-PSDB; AAU69742.

XX Recombinant chimeric protein, useful for detecting and quantifying DNA
PT mutations, e.g. in disease diagnosis, comprises mutation-binding
PT protein and nuclease -

XX PS Disclosure: Page 65; 128pp; English.

XX Sequences AA563224-AA563305 represent cDNA encoding and PCR primers used
CC to amplify cDNA encoding proteins which can be used in the synthesis of
CC chimeric proteins comprising a DNA mutation-binding protein, a linker and
CC a nuclease, by recombinant technology. The chimeric proteins are useful
CC for detection, quantification and mapping of DNA sequence variations
CC including mutations, for example, caused by damage and mismatches. The
CC proteins are able to bind to the site of the DNA mutation and cut it out
CC of the molecule. This is useful for early diagnosis of cancer and other
CC diseases. The proteins used in the invention include human XPF (or
CC ERCC4), human xeroderma pigmentosum complementation groups A, C and E
CC (XPA, XPC and XPE), human Muts homolog 2 (hMSH2), Serratia marcescens
CC nuclease (Nuc), Thermus thermophilus Muts, Escherichia coli Fapy-DNA
CC glycosylase (Fpg), uracil DNA glycosylase (ung), A/G-specific adenine
CC glycosylase (MutY), synthetic T4 endonuclease V (T4 endo V), thymine
CC DNA-glycosylase (TDG), E. coli Uvr A, B and C, and E. coli endonucleases
CC and exonucleases.

XX Sequence 1101 BP; 324 A; 230 C; 246 G; 301 T; 0 other;

Query Match

28.5%; Score 1101; DB 22; length 1101;

Best Local Similarity 100.0%; Pred. No. 1,5e-271;
Matches 1101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 46 ATGGCCCGCGTGTGAGTAGACAGACAGCTGTCTGTGAAGTGTCTGACACTGACGG 105
Db 1 ATGGCCCGCGTGTGAGTAGACAGACAGCTGTCTGTGAAGTGTCTGACACTGACGG 60
QY 106 CTAGTACTGTGCG 165
Db 61 CTAGTACTGTGCG 120
QY 166 CACTGCCACCCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 225
Db 121 CACTGCCACCCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 180
QY 226 TTTATCATCAGCTGAAGATAGAAAGATTGAACACCTCCCTCCGCGTGAACAAATGA 285
Db 181 TTTATCATCAGCTGAAGATAGAAAGATTGAACACCTCCCTCCGCGTGAACAAATGA 240
QY 286 ATCACAAGCAAGTGTGCTATGAAGTTTACACCAAGTGTGTTATTTGGGACAACT 345
Db 241 ATCACAAGCAAGTGTGCTATGAAGTTTACACCAAGTGTGTTATTTGGGACAACT 300
QY 346 AGGATCTTGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 405
Db 301 AGGATCTTGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 360
QY 406 TTGCTGTATAGAGCCCAAGCAATATGAGCTTGTGCAAGCAATTCATCTGCGCTC 465
Db 361 TTGCTGTATAGAGCCCAAGCAATATGAGCTTGTGCAAGCAATTCATCTGCGCTC 420
QY 466 TTTCGCGAGAAAACAAGTGTGTTTATTAAGCTTTCACACAAATGCTGTGCTT 525
Db 421 TTTCGCGAGAAAACAAGTGTGTTTATTAAGCTTTCACACAAATGCTGTGCTT 480
QY 526 GATACGTGTTTTGTGATGTGAAAGATGATGAAATCTTTTGTGAGAAACTGTAT 585
Db 481 GATACGTGTTTTGTGATGTGAAAGATGATGAAATCTTTTGTGAGAAACTGTAT 540
QY 586 CTGTGGCCAAAGTTCATGATGACAGTAATCTATTTTGAACAGCAAACTGAAAGT 645
Db 541 CTGTGGCCAAAGTTCATGATGACAGTAATCTATTTTGAACAGCAAACTGAAAGT 600
QY 646 GTAGAAATCATGTTTCTATGACACCTACCATCTTGTCTATACAGACTGCTATACG 705
Db 601 GTAGAAATCATGTTTCTATGACACCTACCATCTTGTCTATACAGACTGCTATACG 660
QY 706 ATTTTAAATGATGTCTAAAGAACTAAATGCTATTAACCATGCTTGAAGTGAAGAT 765
Db 661 ATTTTAAATGATGTCTAAAGAACTAAATGCTATTAACCATGCTTGAAGTGAAGAT 720
QY 766 TTTATCTTTAGAAAATGCTATTTGAAAACCTTTGACAAAGCAATCCGCCATTTCTG 825
Db 721 TTTATCTTTAGAAAATGCTATTTGAAAACCTTTGACAAAGCAATCCGCCATTTCTG 780
QY 826 CTTTGTGGCAGCAGCTGTGAGCCAAAGCTAAATCTTTATGAGATTTGAAGATTTA 885
Db 781 CTTTGTGGCAGCAGCTGTGAGCCAAAGCTAAATCTTTATGAGATTTGAAGATTTA 840
QY 886 CGAATCTTGTGAGATGCTCTGATGATGATGATGATGATGATGATGATGATGATG 945
Db 841 CGAATCTTGTGAGATGCTCTGATGATGATGATGATGATGATGATGATGATGATG 900
QY 946 TCTCTGAGACCAAGCAAAAGCTTTTGTGCAAAATTCAGTGTGCTTCTGATGCTC 1005
Db 901 TCTCTGAGACCAAGCAAAAGCTTTTGTGCAAAATTCAGTGTGCTTCTGATGCTC 960
QY 1006 AGCAGCTGATGTTTATTAATGCTGAGCAAGGTTTATGATCTTCAGATGCAAAATG 1065
Db 961 AGCAGCTGATGTTTATTAATGCTGAGCAAGGTTTATGATCTTCAGATGCAAAATG 1020
QY 1066 AGTAAAAAGAAAAATATCTGAAAAAATGAAATTAAGAGGCAAGAAAAAAG 1125
Db 1125 AGTAAAAAGAAAAATATCTGAAAAAATGAAATTAAGAGGCAAGAAAAAAG 1080

DB 1021 AGTAAAAAGAAAAATTCGAAAAATTAAGAGGGGAAGAAAAAAG 1080

QY 1126 GAACGTGCTCAGAAAGCAAC 1146
|||||

DB 1081 GAACGTGCTCAGAAAGCAAC 1101

RESULT 8

ABL87494/c
ID ABL87494 standard; cDNA; 601 BP.
XX
AC ABL87494;
XX
DT 17-MAY-2002 (first entry)
XX
DE Human ovarian cancer related cDNA clone SEQ ID NO:10472.
XX
KW Human; ovarian cancer; ovarian tumour; cytostatic; gene; ss.
XX
OS Homo sapiens.
XX
PN WO200192581-A2.
XX
PD 06-DEC-2001.
XX
PF 29-MAY-2001; 2001WO-US17756.
XX
PR 26-MAY-2000; 2000US-207484P.
XX
PA (CORI-) CORIXA CORP.
XX
PI Algate PA, Harlocker SL, Jones R;
XX
DR WPI: 2002-122075/16.
XX
PT Composition for therapy and diagnosis of ovarian cancer comprising
PT polypeptide of a ovarian tumor polypeptide, polynucleotide encoding
PT polypeptide, antibody specific to polypeptide or T cell expressing
PT polypeptide -

PS Claim 1; SEQ ID 10472; 489pp; English.

CC The present invention describes a composition (I) comprising: carriers
CC and immunostimulants; and a polypeptide (II) of a ovarian tumor
CC polypeptide encoded by a polynucleotide (III) having a cDNA sequence
CC (S1) from the 10912 nucleotide sequences as given in ABL77003 to
CC ABL87934, (III) encoding (II) having a sequence (S2), a T cell
CC population of (II), or antigen presenting cells that express (II).
CC (I) has cytostatic activity. An oligonucleotide (IV) that hybridises to
CC (S1) can be used for detecting ovarian cancer in a patient's biological
CC sample preferably serum or ovarian tissue. The method comprises
CC contacting a biological sample from a patient with (IV), detecting the
CC amount of polynucleotide hybridising to (IV) and comparing the amount to
CC a predetermined cutoff value and thereby detecting ovarian cancer in the
CC patient, where the amount of polynucleotide hybridising to (IV) is
CC detected preferably by polymerase chain reaction (PCR). (I) comprising
CC (III) and/or (II) is useful for stimulating and/or expanding T cells
CC specific for an ovarian tumor protein comprising contacting T cells
CC with (III) or (II). (III) is useful in design and preparation of
CC ribozyme molecules for inhibiting expression of the tumor polypeptides
CC and proteins in tumour cells; and to isolate a full length gene from a
CC suitable library e.g., a tumour cDNA library using well known
CC techniques.

XX
XX Sequence 601 BP; 180 A; 134 C; 118 G; 166 T; 3 other.

QY Query Match 6.6%; Score 256.2; DB 24; Length 601;
Best Local Similarity 97.7%; Pred. No. 2.8e-55;
Matches 258; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1561 GAATTAAGCAGTACCCAGAAAGTCCCGGAGAAATTAGCATGGAAGATTGGATGA 1620
|||||

DB 601 GAATTAAGCAGTACCCAGAAAGTCCCGGAGAAATTAGCATGGAAGATTGGATGA 542

QY 1621 AATTGTCATGCGATGCTGCTTTGGAATCCGAAAGAACCCCTCATTCATCCGCG 1680
|||||

DB 541 AATNTGTCATGCGATGCTGCTTTGGAATCCGAAAGAACCCCTCATTCATCCATCCG 482
|||||

QY 1681 CTCTGGGTTGACGCGACCCCTATGCTCTGACAAAGGCTACTACATGAAGTGGAGCCAAGA 1740
|||||

DB 481 CTCTGGGTTGACGCGACCCCTATGCTCTGACAAAGGCTACTACATGAAGTGGAGCCAAGA 422
|||||

QY 1741 TACGTGTTCTTTATGACGCGACGCTTGTGCGGACGCTTGAATTTACAGGCGC 1800
|||||

DB 421 TACGTGTTCTTTATGACGCGACGCTTGTGCGGACGCTTGAATTTACAGGCGC 362
|||||

QY 1801 AGTAGCCTGGGAAACCTCTGAGG 1824
|||||

DB 361 AGTAGCCTGGGAAACCTCTGAGG 338
|||||

RESULT 9

ABL06559
ID ABL06559 standard; cDNA; 2995 BP.
XX
AC ABL06559;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 14159.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ss.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
XX
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI: 2001-656860/75.
XX
DR P-PSDB: ABB62456.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -

PS Claim 1; SEQ ID NO 14159; 21pp + Sequence Listing; English.

XX
XX The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signalling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
XX sequences (ABL01840-ABL16175) and the encoded proteins
XX (ABB57737-ABB72072).

CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at [ftp.wipo.int/pub/published_pcl_sequences](http://wipo.int/pub/published_pcl_sequences).

XX
XX Sequence 2995 BP; 775 A; 749 C; 807 G; 664 T; 0 other.

QY Query Match 6.6%; Score 253.2; DB 23; Length 2995;
Best Local Similarity 46.9%; Pred. No. 3.7e-54;
Matches 1270; Conservative 0; Mismatches 1303; Indels 135; Gaps 10;

Db	2293	ATTGTGTTGGAGGAAAGTCCATTTCGTGATTTGATTTGGGTCCTGAAATTCGGGACATTTG	2352
Qy	2236	TACAGCGATGCATCTCCATGTCCCGCTACTACAGCGTCGCCGTCTTGATTGATTTG	2295
Db	2353	TTCAACCAATGGGTGCMAATGAGAGGGGCATTTATGCAAAACCCCATCTTTTGATTGATTC	2412
Qy	2296	GACCCTAGCAACCTTTTCTCTCTCACTTCCGAGGTGCCCTGTGTTCAGAG-----ATC	2349
Db	2413	GATCAGAAATTAACCCCTTTCACCTGCAGGGCAATTTATGTATGCGACGACACAGATG	2472
Qy	2350	TTCCAAATGACATTAAGTTCCAAACTCACTCTTTTACACTTTCACCTTCCACAGACTACGG	2409
Db	2473	GCCAAATGCAGATATTGTGCGAAGAACTGCAACTCTTACCTTACATTTTCCCAAGCTGCGA	2532
Qy	2410	ATTTCTGTGGTCCCTCTCTCTATATGCAAGGGGAGTTGTTGAGAGGTGAACAAGC	2469
Db	2533	CTCATTTGGTCAACCCAGTCCGATGTCACACTGCCCCACCTATTTGAGGAGCTGGAAGCTGGT	2592
Qy	2470	AAGCCACAGCCCTGATGCGCGCGACAGCACTGGCCATTACAGAGATTTCTAAACCTTCCC	2529
Db	2593	AAACCGGAACCGATTCCTCAAAACGGGGCGCCCTTGGGAGAGGACGACCAACGGGGGT	2652
Qy	2530	GAGTCAGAGAGACTATATATCTCTGTCCCAAGACTTCTTTAAAAATGCCAGGGGTGAAT	2589
Db	2653	GAGCAATATGCATTTTCAACAGCGGCATCTGCAATTTCTCTGCTACGTGCCCCGGCGTTTCAAT	2712
Qy	2590	GCCAAACACTGGCGCTCTTGTATGCAACCCGTTAAGAAATGCGCAATTAAGCAGCCCTC	2649
Db	2713	ACGGCAATATATCAATGCGCTACTCCGGAAGGGCGGACGCTCGGCAATTTGTCTGCGC	2772
Qy	2650	TCACAAGACGAGCTCAGCAGATATCTTGGGGAATGCTGCAAAATGCCAAAGACTTTATGAT	2709
Db	2773	AGCCCAAGGAAGACTGGAGAGGCTGCTGCACTGCGACGAGAGAGGCGCCAAAGCTGTACGAT	2832
Qy	2710	TTTCATTCA 2717	
Db	2833	ATCTCTGCA 2840	
RESULT 10			
ABL06558	ID	ABL06558 standard; cDNA; 5313 BP.	
AC	ABL06558;		
XX			
DT	26-MAR-2002 (first entry)		
XX			
DE	Drosophila melanogaster expressed polynucleotide SHQ ID NO 14156.		
XX			
KW	Drosophila; developmental biology; cell signalling; insecticide;		
KW	pharmaceutical; gene; ss.		
OS	Drosophila melanogaster.		
XX			
PN	WO20011042-A2.		
XX			
PD	27-SEP-2001.		
XX			
PF	23-MAR-2001; 2001WO-US09231.		
XX			
PR	23-MAR-2000; 2000US-191637P.		
PR	11-JUL-2000; 2000US-0614150.		
XX			
PA	(PEKE) PE CORP NY.		
XX			
P1	Venter JC, Adams M, Li PWD, Myers EW;		
XX			
DR	WPI; 2001-656860/75.		
DR	P-PSDB; ABB62455.		
XX			
PT	New isolated nucleic acid detection reagent for detecting 1000 or more		
PT	genes from Drosophila and for elucidating cell signalling and cell-cell		
PT	interactions -		

XX	Claim 1; SEQ ID NO 14156; 21pp + Sequence Listing; English.
PS	
XX	The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from <i>Drosophila</i> . The invention is useful in developmental biology and in elucidating cell signalling and cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-ABB7202).
CC	(ABB57737-ABB7202).
CC	The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.
CC	
XX	
XX	Sequence 5313 BP; 1466 A; 1195 C; 1267 G; 1385 T; 0 other;
SO	
	Query Match 4.1%; Score 159; DB 235; Length 5313; Best Local Similarity 53.5%; Pred. No. 6; 8e-30; Matches 358; Conservative 0; Mismatches 305; Indels 6; Gaps 11;
OY	2055 GCAAAGCATAGTTGTGGATATCGTGAAATTTCGAAGTGCCTTCATCTGATCCATCG 2114
DB	3490 GCCAAAAGGATGTGGAGATATCGTAGTTGGCTCAGATTCGCCCTTTATTCACAA 3549
OY	2115 TCGGGGCAATGATGACATTGAACCGGTACTTTAAGGTGGAGATTACATCCATCCACA 2174
DB	3550 GCGTGTCTGGAAAGTGTCTGCCCTTGACCATTAACGATTGGCGAATTATATCTCACCGGA 3609
OY	2175 AATGTGCGTGGAGCGCCAGACGATGATCAGTATTAATCGGCTTTTAAATAACGGCGCCT 2234
DB	3610 TATTGTGTGGAGCGAAAGTCATTTCTGATTGTGATTGGCTCTCGAATTGGCGAGATT 3669
OY	2235 CTACAGCCAGTGCATCTCATGTCCCGCTACATACAGCGTCCCCGTCCTCTATGAGTT 2294
DB	3670 GTACCAACCATCCGTGCAAAATCAGCGGGGATTATGCAAAAGCCCATTCTTTATTTAGAGTT 3729
OY	2295 TCACCTTAGCAGAAGCTTTCTCTCTCACTCCGAGGTGCTTGTTCAGAG-----AT 2348
DB	3730 CGATCAGATTAAGCCCTTTCACCTTCGAGGGCAAATTTATGTTATGSCAACACAGCAT 3789
OY	2349 CTCACAGCAATGACATTAGTCCAAACTACTCTTTACATCTTACCTTCCCAAGCTACG 2408
DB	3790 GGCCAATGCAGATTATGTGACAGAAATCGCAACTTCTTAGCTTACATTTTCCAAAGCTGCG 3849
OY	2409 GAATTCCTGTGATCCCTCTCTCATGCAAGCGCGGAGTGTGTGAGAGCTGAACAAG 2468
DB	3850 ACTCATTTTGTATACCCAGTCCGTATGBCCATCTGCCACGATATTGCGAGAGCTTAAGCTGGG 3909
OY	2469 CAAGCCACAGCTGATGTGCGGCGACAGCACTGCGCATTAACAGAGATTCTGAACCCCTTCC 2528
DB	3910 TTAACCCGGAACCGGATTCCTCAAAAGCGGCGCCCTTGGGACAGCAGACCAAGCGGGG 3969
OY	2529 CGAGTCAGAGAACTTAATTCCTGTGTCGCCAAGACTTCTGTAAAAAATGCCAGGGGTGA 2588
DB	3970 TGAGCAATTGCAATTTCACAGCGGAGCTACGATTTCTGCTACGTTCTGCCGGGGTTCA 4029
OY	2589 TCCCAAAAACCTCGCTCTCTTGATGATCACCACACCTTTAAGACATCGCAGATTTAGCAGCCT 2648
DB	4030 TACGGCGAATTATACATAGGCTACTCCGGAAGGGCGGACGTCGCGCATTTGTTGCTGGG 4089
OY	2649 GTACACAGACGAGCTACAGATATTCTGGGGAATGCTGCATAATGCCAAGACCTTTATGA 2708
DB	4090 CAGCGAAGAGAACTGGAGAGCTGCTGCAGTGCGAGAGAGACGCCAAGCTCTGTACGA 4149
OY	2709 TTTCATTCA 2717
DB	4150 TATCTCGCA 4158
RESULT 11	
ID	ABK35494/C ABK35494 standard: DNA: 3857 BP.

XX 01-SEP-2000: 2000DE-1043826.
PR 05-SEP-2000: 2000DE-1044543.
XX
XX (EPiG-) EPIGENOMICS AG.
PA Olek A, Piepenbrock C, Berlin K, Guetig D;
XX
XX WPI: 2002-371829/40.
XX
XX Determining the degree of cytosine methylation in genomic DNA, useful
PT for diagnosis and prognosis, comprises selective hybridization of
PT amplicons from chemically treated DNA
XX
XX Claim 12: 56pp + Sequence Listing; 56pp; German.
PS
XX This invention describes a novel method for determining the degree of
CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC genomic sample of DNA. The sample is treated chemically to convert
CC cytosine (C) but not methylated C, to uracil, then part of the genomic
CC DNA that contains the target C is amplified to form a labeled amplicon.
CC The amplicon is hybridised to two classes, each with at least one
CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
CC and the degree of hybridisation to both classes is determined from the
CC label on the amplicon. From the ratio of labels hybridised to the two
CC classes (1) for diagnosis and/or prognosis of side effects of
CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
CC of the central nervous, cardiovascular, gastrointestinal and respiratory
CC systems etc., particularly by detecting mutations or single nucleotide
CC polymorphisms (SNP's); and (11) for differentiation of cell or tissue
CC types and for investigating cell differentiation. The method allows the
CC methylation status of many C residues to be determined simultaneously.
CC AB03410-AB054121 represent genomic DNA sequences used to illustrate the
CC method for determining the degree of cytosine methylation described in
CC the disclosure of the invention.
XX
XX Sequence 574 BP: 79 A: 54 C: 173 G: 268 T: 0 other;
SQ
Query Match 3.5%; Score 135.6; DB 24; Length 574;
Best Local Similarity 75.7%; Pred. No. 2.1e-24;
Matches 168; Conservative 0; Mismatches 54; Indels 0; Gaps 0;
QY 1 GGAAGACCTTCATGAGTACAGCGGCGGCTCGACGATTCGATGCGCGCGCTGCTG 60
DB 343 GGAAGAGTTTATGAGTATGAGTATGAGTATGAGTATGAGTATGAGTATGAGTATG 402
QY 61 GAGTACGAGGACGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 120
DB 403 GAGTACGAGGACGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 462
QY 121 CGCGGGCTGCGCGGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 180
DB 463 CGCGGGCTGCGCGGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 522
QY 181 TGCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 222
DB 523 TGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 564

RESULT 15
ABQ21889/c
ID ABQ21889 standard; DNA: 574 BP.
XX
AC ABQ21889;
XX
XX 12-JUL-2002 (first entry)
XX
DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 8480.
XX
XX Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
KW drug; side effect; cancer; central nervous system; cardiovascular;
KW gastrointestinal; respiratory system; single nucleotide polymorphism;

KW SNP; cell differentiation; ds.
XX
XX Homo sapiens.
OS
XX MO200218632-A2.
PN
XX
XX 07-MAR-2002.
PD
XX
XX 01-SEP-2001; 2001WO-EP10074.
PF
XX
XX 01-SEP-2000: 2000DE-1043826.
PR 05-SEP-2000: 2000DE-1044543.
XX
XX (EPiG-) EPIGENOMICS AG.
PA Olek A, Piepenbrock C, Berlin K, Guetig D;
XX
XX WPI: 2002-371829/40.
XX
XX Determining the degree of cytosine methylation in genomic DNA, useful
PT for diagnosis and prognosis, comprises selective hybridization of
PT amplicons from chemically treated DNA
XX
XX Claim 12: 56pp + Sequence Listing; 56pp; German.
PS
XX This invention describes a novel method for determining the degree of
CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC genomic sample of DNA. The sample is treated chemically to convert
CC cytosine (C) but not methylated C, to uracil, then part of the genomic
CC DNA that contains the target C is amplified to form a labeled amplicon.
CC The amplicon is hybridised to two classes, each with at least one
CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
CC and the degree of hybridisation to both classes is determined from the
CC label on the amplicon. From the ratio of labels hybridised to the two
CC classes (1) for diagnosis and/or prognosis of side effects of
CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
CC of the central nervous, cardiovascular, gastrointestinal and respiratory
CC systems etc., particularly by detecting mutations or single nucleotide
CC polymorphisms (SNP's); and (11) for differentiation of cell or tissue
CC types and for investigating cell differentiation. The method allows the
CC methylation status of many C residues to be determined simultaneously.
CC AB03410-AB054121 represent genomic DNA sequences used to illustrate the
CC method for determining the degree of cytosine methylation described in
CC the disclosure of the invention.
XX
XX Sequence 574 BP: 268 A: 173 C: 54 G: 79 T: 0 other;
SQ
Query Match 3.5%; Score 135.6; DB 24; Length 574;
Best Local Similarity 75.7%; Pred. No. 2.1e-24;
Matches 168; Conservative 0; Mismatches 54; Indels 0; Gaps 0;
QY 1 GGAAGACCTTCATGAGTACAGCGGCGGCTCGACGATTCGATGCGCGCGCTGCTG 60
DB 232 GGAAGAGTTTATGAGTATGAGTATGAGTATGAGTATGAGTATGAGTATGAGTATG 173
QY 61 GAGTACGAGGACGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 120
DB 172 GAGTACGAGGACGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 113
QY 121 CGCGGGCTGCGCGGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 180
DB 112 CGCGGGCTGCGCGGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 53
QY 181 TGCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 222
DB 52 TGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 11

Search completed: November 26, 2002, 20:11:50
Job time : 1054.78 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 26, 2002, 19:45:25 ; Search time 3612.58 Seconds
(without alignments)
17291.221 Million cell updates/sec

Title: US-09-919-497-14
Perfect score: 3857
Sequence: 1 ggaagagctccatgagtc.....agaatacaactaagctca 3857

Scoring table: IDENTITY:NUC
Gapop 10.0, Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues
Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: em_estba:*
2: em_esthum:*
3: em_estln:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hlc:*
9: gb_est1:*
10: gb_est2:*
11: gb_hlc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rtd:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	1062.2	27.5	1083	11 BC020741
2	780.6	20.2	818	13 B1522552
3	671.8	17.4	748	12 BG620282
4	653.8	17.0	741	12 BG181154
5	610.4	15.8	1119	11 AK018375
6	537.8	13.9	344	14 BQ013114

C 7	537.4	13.9	581	10 AW977575	AW977575 EST898684
C 8	530.6	13.8	683	10 AW242081	AW242081 x021c12.x
C 9	520.6	13.5	530	14 BQ011470	BQ011470 UT-1-BC1P
C 10	513.6	13.3	829	13 B1555512	B1555512 603239318
C 11	499.6	13.0	515	14 BQ310815	BQ310815 OV3-BN004
C 12	499.6	12.9	524	13 BM671280	BM671280 UT-E-CQ1-
C 13	485.6	12.6	923	14 BE818393	BE818393 RC5-BN028
C 14	482.2	12.5	618	14 BQ944411	BQ944411 AGENCOURT
C 15	467	12.1	596	10 AV685090	AV685090 AV685090
C 16	466	12.1	466	14 BM710111	BM710111 UT-E-CQ1-
C 17	457.4	11.9	459	9 AL705565	AL705565 DKFZP6861
C 18	456.2	11.8	509	12 BE837466	BE837466 RC2-FN009
C 19	453.6	11.8	460	9 AT025538	AT025538 ov76d11.x
C 20	452.2	11.7	752	12 BG724387	BG724387 602692867
C 21	452	11.7	464	9 AA256859	AA256859 zr83a11.r
C 22	451.4	11.7	651	10 BB640850	BB640850 BB640850
C 23	447.4	11.6	708	9 AJ447405	AJ447405 AJ447405
C 24	437	11.3	449	9 AA291199	AA291199 z547c05.r
C 25	436.2	11.3	463	9 AT018046	AT018046 ov64c11.s
C 26	433.6	11.2	641	9 AJ447403	AJ447403 AJ447403
C 27	422.2	10.9	887	10 BE565954	BE565954 601328232
C 28	421.4	10.9	434	9 AA770518	AA770518 ah74d10.s
C 29	416.6	10.8	640	12 BG866533	BG866533 602785491
C 30	407.8	10.6	418	9 AT431784	AT431784 t125b12.x
C 31	400.8	10.4	715	13 B1459712	B1459712 603200751
C 32	395.6	10.3	425	9 AA774566	AA774566 a127c07.s
C 33	385.6	10.0	403	10 AV692790	AV692790 AV692790
C 34	383.2	9.9	801	9 AJ447404	AJ447404 AJ447404
C 35	380.4	9.9	654	9 AJ447402	AJ447402 AJ447402
C 36	358.8	9.3	374	12 BE818447	BE818447 RC5-BN028
C 37	355	9.2	523	10 AW271424	AW271424 x514d04.x
C 38	340.8	8.8	471	10 BB856543	BB856543 BB856543
C 39	339.6	8.8	357	9 AA638976	AA638976 ns01b08.s
C 40	322	8.3	583	13 B1693586	B1693586 603341969
C 41	312.8	8.1	668	10 AW921189	AW921189 EST352493
C 42	309.4	8.0	501	9 AA436963	AA436963 zv72f08.r
C 43	296	7.7	296	9 AA255461	AA255461 zr83a11.s
C 44	295.2	7.7	312	9 AA723776	AA723776 ah62f10.s
C 45	282	7.3	470	9 AT013171	AT013171 EST207846

ALIGNMENTS

RESULT 1
LOCUS BC020741 1083 bp mRNA linear HTC 22-JAN-2002
DEFINITION Homo sapiens, Similar to excision repair cross-complementing rodent repair deficiency, complementation group 4, clone IMAGE:4732192, mRNA.
ACCESSION BC020741
VERSION BC020741.1 GI:18089125
KEYWORDS HTC.
SOURCE Homo sapiens.
ORGANISM Homo sapiens.
REFERENCE 1 (bases 1 to 1083)
AUTHORS Strausberg, R.
TITLE Direct Submission
JOURNAL Submitted (03-JAN-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
REMARK NIH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: cgapbs-remail.nih.gov
Tissue Procurement: CLONTECH
CDNA Library Preparation: CLONTECH Laboratories, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILMN)
DNA Sequencing by: Sequencing Group at the Stanford Human Genome Center, Stanford University School of Medicine, Stanford, CA 94305
Web site: <http://www.shgc.stanford.edu>

Contact: (Dickson, Mark) mcdexall.stanford.edu
 Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,
 R. M.

Clone distribution: MGC clone distribution information can be found
 through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 Series: IRML Plate: 37 Row: h Column: 11
 This clone has the following problem: frame shifted.

FEATURES

source
 1. 1083
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4732192"
 /issue_type="Placenta"
 /clone_lib="NIH_MGC_79"
 /lab_host="DH10B"
 /note="vector: pDNR-LIB"
 BASE COUNT 322 a 229 c 235 g 297 t
 ORIGIN

Query Match 27.5%; Score 1062.2; DB 11; Length 1083;

Best Local Similarity 98.8%; Pred. No. 9.3e-214;
 Matches 1070; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

33 TCGACGATTTCCATGCGCCGCTGCTGAGTACAGACGACGCTGCTGAGTACTGCT 92
 1 TCGACGATTTCCATGCGCCGCTGCTGAGTACAGACGACGCTGCTGAGTACTGCT 60
 93 CGACATGACGGGCTAGTGTGCGCCGCGGCGCTGCGGCGGACGCGCTCTCTACCA 152
 61 CGACATGACGGGCTAGTGTGCGCCGCGGCGGCGGCGGCGGCGGCTCTCTACCA 120
 153 CTTTTCACAGCTGACCTGCGCCAGCCAGCTGCTGCTGCTGCTGCTGCTGCTGCT 212
 121 CTTTTCACAGCTGACCTGCGCCAGCCAGCTGCTGCTGCTGCTGCTGCTGCTGCT 180
 213 CGAGAGAGATTTTATCAATCAGCTGAGATGAGAGAGATTTTATCAATCAGCTG 272
 181 CGAGAGAGATTTTATCAATCAGCTGAGATGAGAGAGATTTTATCAATCAGCTG 240
 273 TGTAAACAATGAAATCACAAGCAAGCTGCTGAGATTTTATCAATCAGCTGCT 332
 241 TGTAAACAATGAAATCACAAGCAAGCTGCTGAGATTTTATCAATCAGCTGCT 300
 333 ATTTCGACAGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 392
 301 ATTTCGACAGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 360
 393 AATTACTGCGATCTTGTGTATGAGAGCCAGCAATATGATGATGATGATGATGAT 452
 361 AATTACTGCGATCTTGTGTATGAGAGCCAGCAATATGATGATGATGATGATGAT 420
 453 CATCTGGCGCTCTTGGCCAGAAAACAGTGGTTTATTAAGCTTTCACAGCAA 512
 421 CATCTGGCGCTCTTGGCCAGAAAACAGTGGTTTATTAAGCTTTCACAGCAA 480
 513 TCGTGTGCTTGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 572
 481 TCGTGTGCTTGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 540
 573 GAGGAACCTATCTGTGCGCAAGTTCATGTAGACGATTAATCTTTTATGACGA 632
 541 GAGGAACCTATCTGTGCGCAAGTTCATGTAGACGATTAATCTTTTATGACGA 600
 633 CAACCTGAGTGTGAGAAATCATGTTTCTATGACACCTACCATGCTCTATACGAC 692
 601 CAACCTGAGTGTGAGAAATCATGTTTCTATGACACCTACCATGCTCTATACGAC 660
 693 TCGTATCTGACATTTTAAATGATGATGATGATGATGATGATGATGATGATGAT 752
 661 TCGTATCTGACATTTTAAATGATGATGATGATGATGATGATGATGATGATGAT 720
 753 TGAAGTGAAGATTTTATCTTTAGAAAATGCTATTGAAAACTTTTGCACAAGCAATCG 812

|||||
 Db 721 TGAAGTGAAGATTTTATCTTTGAAAAATGCTATTGGAAAACTTTTGCACAAGCAATCG 780
 813 CCATTATCTGAGATCTTTTGGCACCAGCTTGAGGCAAGCAATATCTTATGTTGAGGA 872
 Db 781 CCATTATCTGAGATCTTTTGGCACCAGCTTGAGGCAAGCAATATCTTATGTTGAGGA 840
 873 TTTGAAGATTTATGAGAACTTTGCTGAGTATCTCTCATGATGATGATGATGATGAT 932
 Db 841 TTTGAAGATTTATGAGAACTTTGCTGAGTATCTCTCATGATGATGATGATGATGAT 900
 933 TATCTTCTGGAATCTCTGAGACCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCA 992
 Db 901 TATCTTCTGGAATCTCTGAGACCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCA 960
 993 GTTTCCTGAGTCTGAGACCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCA 1052
 Db 961 GTTTCCTGAGTCTGAGACCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCA 1020
 1053 AGATGCCAAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1112
 Db 1021 AGATGCCAAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1080
 1113 AGA 1115
 1081 AAA 1083
 RESULT 2
 B1522552 818 bp mRNA linear EST 29-AUG-2001
 LOCUS 603175243F1 NIH_MGC_121 Homo sapiens cDNA clone IMAGE:5239666 5',
 DEFINITION mRNA sequence.
 ACCESSION B1522552
 VERSION B1522552.1 GI:15347344
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 818)
 NIH-MGC <http://mgc.nci.nih.gov/>
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 CONTACT Robert Strausberg, Ph.D.
 Email: cgabs-remail.nih.gov
 Tissue Procurement: Life Technologies, Inc.
 CDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: InCyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: LLML1604 row: e column: 11
 High quality sequence step: 802.
 Location/Qualifiers
 1. 818
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:5239666"
 /clone_lib="NIH_MGC_121"
 /lab_host="DH10B"
 /note="Organ: Brain; Vector: pCMV-Sport6; Site_1: NotI;
 Site_2: EcoRV (destroyed); RNA source anonymous pool of 3
 fetal brains, female age 20 weeks, female age 24 weeks,
 and male age 26 weeks. Library is oligo-dT primed and
 directionally cloned (EcoRV site is destroyed upon
 cloning). Average insert size 1.7 kb, insert size range
 0.7-3.5 kb. Library is normalized and enriched for
 full-length clones and was constructed by C. Gruber
 (Invitrogen). Research Genetics tracking code 017. Note:
 this is a NIH_MGC Library."
 BASE COUNT 241 a 158 c 167 g 252 t

QY 513 TGGTGTGCTTTGATACGTGTTTGTTCATGTGGAAAGATGATGAAATCTTTTGT 572
 DB 482 TGGGTGCTGCTTTGATACGTGTTTGTTCATGTGGAAAGATGATGAAATC-TTTTGT 540
 QY 573 GAGGAACCTATCTGTGGCCAAAGTTCCATGTAGCAGTAACTATTTTGAACAGCA 632
 DB 541 GAGGAACCTATCTGTGGCCAAAGTTCCATGTAGCAGTAACTATTTTGAACAGCA 600
 QY 633 CAACCTGAAGTGTAGAAATCCATGTTCTATGACACCTACCATGCTGTATACAGAC 692
 DB 601 CAACCTGAAGTGTAGAAATCCATGTTCTATGACACCTACCATGCTGTATACAGAC 660
 QY 693 TGGTATCTGACATTTTAAATGATGCTTAAAGAACTAAATGCCATACCCATGCT 752
 DB 661 TGTATCTGTGACATTTTAAATGATGCTTAAAGAACTAAATGCCATACCCATGCT-- 718
 QY 753 TGAAGTGGAGATTTATCTTTAGAAAATG 781
 DB 719 TTGAGGTGCACGATTTATCTTTAGAAAAG 747

RESULT 4
 LOCUS BG181154 741 bp mRNA linear EST 06-FEB-2001
 DEFINITION 602329292F1 NIH_MGC_91 Homo sapiens cDNA clone IMAGE:4430639 5',
 mRNA sequence.
 ACCESSION BG181154
 VERSION BG181154.1 GI:12687857
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 AUTHORS NIH-MGC http://mgc.nci.nih.gov/
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgrabbs-remail.nih.gov
 Tissue Procurement: DCTP/DRP

CDNA Library Preparation: Life Technologies, Inc.
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LLNL0184 row: g column: 24
 High quality sequence stop: 677.
 Location/Qualifiers

FEATURES

source
 1..741
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4430639"
 /clone_lib="NIH_MGC_91"
 /tissue_type="adenoecarcinoma, cell line"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: Prostate; Vector: PCMV-SPORT6; Site:1; Ncti:
 Site:2; Salt: Cloned unidirectionally; oligo-dr primed.
 Average insert size 1.4 kb. Library enriched for
 full-length clones and constructed by Life Technologies.
 Note: this is a NIH_MGC Library."
 BASE COUNT 218 a 151 c 150 g 222 t
 ORIGIN

Query Match 17.0%; Score 653.8; DB 12; Length 741;
 Best Local Similarity 96.9%; Pred.No.1.5e-127;
 Matches 720; Conservative 0; Mismatches 17; Indels 6; Gaps 5;
 QY 248 AAGGAGTTGAACACCTCCCTGCCGTGTAAACAATGAAATCACAAGCAGTGGCTATG 307
 DB 1 AAGGAGTTGAACACCTCCCTGCCGTGTAAACAATGAAATCACAAGCAGTGGCTATG 60

QY 308 AAGTTTACACAAGAGTGTGTTATATTTCGACAGATAGATCTTGTGACTTCT 367
 DB 61 AAGTTTACACAAGAGTGTGTTATATTTCGACAGATAGATCTTGTGACTTCT 120
 QY 368 TGAAGTATAGAAATCTTCAAGATTTAACTAGGATCTTGTGTATAGAGCCACAGAA 427
 DB 121 TGAAGTATAGAAATCTTCAAGATTTAACTAGGATCTTGTGTATAGAGCCACAGAA 180
 QY 428 TAAATGAGCTTGTCAAGAAAGCAATCTTGGGCGCTTTCGCCGCAAAAACAAACGCTG 487
 DB 181 TAAATGAGCTTGTCAAGAAAGCAATCTTGGGCGCTTTCGCCGCAAAAACAAACGCTG 240
 QY 488 GTTTTATTAAGCTTTTCAACAGAAATGCTTGTGCTTGTGATGCTGTTTTGTGATGCTG 547
 DB 241 GTTTTATTAAGCTTTTCAACAGAAATGCTTGTGCTTGTGATGCTGTTTTGTGATGCTG 300
 QY 548 AAGAGTATAGAAATCTTGTGTGAGGAAGTATCTGTGGCCA-GGTTCCATGTA 606
 DB 301 AAGAGTATAGAAATC-TTTTGTGAGGAAGTATCTGTGGCCAAGGTTCCATGTA 359
 QY 607 GCAGTAACTCATTTTGAACAGCAACCAACCTGAAGTTGTAGAAATCCATGTTTCTATG 666
 DB 360 GCAGTAACTCATTTTGAACAGCAACCAACCTGAAGTTGTAGAAATCCATGTTTCTATG 419
 QY 667 ACACCTACCAATGCTTGTGATACAGAGCTGATCTGACATTTTAAATGCAATGCTTAAAG 726
 DB 420 ACACCTACCAATGCTTGTGATACAGAGCTGATCTGACATTTTAAATGCAATGCTTAAAG 479
 QY 727 GAACCTAAATGCCATTAACCCATGCTTGAAGTGAAGATTTATCTTTAGAAAATGCTATT 786
 DB 480 GAACCTAAATGCCATTAACCCATGCTTGAAGTGAAGATTTATCTTTAGAAAATGCTATT 539
 QY 787 GGAACCTTTTGAACAAAGCAATCCGCCATTTCTGATGCTTGTGCGCCAGCCTTGA 846
 DB 540 GGAACCTTTTGAACAAAGCAATCCGCCATTTCTGATGCTTGTGCGCCAGCCTTGA 598
 QY 847 GCCAAGACTAAATCTTGTAGTTCAGAG-TTTGAAGATATTGCAACTTGTGTCAGATATCT 905
 DB 599 GCCAAGACTAAATCTTGTAGTTCAGAG-TTTGAAGATATTGCAACTTGTGTCAGATATCT 658
 QY 906 CTCTCAGTATGATTTGTCTACAT--TTCTTAATCTTGTGAATCTGTGAGAGCAGGAA 963
 DB 659 CTCTCAGTATGATTTGTCTACATCTTCTTAATCTTGTGAATCTGTGAGAGCAGGAA 718
 QY 964 AAGCTTTTGTGTCAGAAATTCAGG 986
 DB 719 CAACGCTTTTGGCAGAAATCCCG 741

RESULT 5
 LOCUS AK018375 1119 bp mRNA linear HTC 19-JAN-2002
 DEFINITION Mus musculus 16 days embryo lung cDNA, RIKEN full-length enriched
 library, clone:8430401P11:excision repair cross-complementing
 rodent repair deficiency, complementation group 4, full insert
 sequence.

ACCESSION AK018375
 VERSION AK018375.1 GI:12858047
 KEYWORDS HTC; CAP trapper.
 SOURCE Mus musculus (strain:C57BL/6J) 16 days embryo lung cDNA to mRNA,
 clone:lib:RIKEN full-length enriched mouse cDNA library
 clone:8430401P11.
 ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
 AUTHORS Carninci, P. and Hayashizaki, Y.
 TITLE High-efficiency full-length cDNA cloning
 JOURNAL Meth. Enzymol. 303, 19-44 (1999)
 MEDLINE 99279253
 PUBMED 10349636
 REFERENCE Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,

Db 368 GATTCTTGACAGTAGAATACCTTCAGATTATATACTGGCATCTCTGCTAGCAGGCA 427
Qy 421 CACAGATATATCGATCTTGTGCAAGAACATTCATCTTGCGCTCTTGGCCAGAAAC 480
Db 428 CACAGATATATGAGTCGCGCAGGAAGCTTCATCTGCGCTCTTGGCCAGAAAC 487
Qy 481 AAACGGTGTCTTATTAAGCTTACAGAACATGCTGTGCTGCTGATACCTGTTTGT 540
Db 488 AACGGCGGTTCTATCAAGCTTTCACCGACACAGCTGTGCTGCTGATACCTGTTTGT 547
Qy 541 CATGTGGAAGAGTGTGAGAAATCTTTTGTGAGGAACCTGATCTGTGGCCAGGTC 600
Db 548 CACGTGGAAGAGTGTGAGGAACCTTTTGTGAGGAACCTGATCTGTGGCCAGGTC 607
Qy 601 CACTGAGCACTAATCATTTTATAGAACAGCAACCAACCTGAGTGTAGAAACCTGTT 660
Db 608 CACTGAGCACTAATCATTTTGTGAGAACACACAGCTGAACTGTAGGATTCACGTC 667
Qy 661 TCTATGACACCTACATGCTGCTATACAGACGCTATCTGAGCACTTTTAAATGATGT 720
Db 668 TCCATGACCTCCGCCATGCTGTGCATTCACAGCGCCATCTAGACATCTTAAATGATGT 727
Qy 721 CTAAAGAACTAAATGCAATACCCATGCTGAAGTGAAGATTTATCTTTAGAAAT 780
Db 728 CTGAAGAACTAAATGCAATGCAACCCGCTGAGTGAAGATTTGTCAGTAAAC 787
Qy 781 GCTATGGAACCTTTTGACAG 804
Db 788 GCTCTTGGAGCCATTCGACAG 811

RESULT 5
B0013114/c 544 bp mRNA linear EST 26-MAR-2002
LOCUS B0013114
DEFINITION UI-1-BC1P-ay1-f-03-0-UI s1 NCI-CGAP_P13 Homo sapiens cDNA clone
B0013114
VERSION B0013114.1 GI:19738015
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 544)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaps-r@mail.nih.gov
Tissue Procurement: Dr. Steven Brown
CDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Clone distribution information can be obtained
from Dr. M. Bento Soares, bento-soares@uiowa.edu
Seq primer: M13 FORWARD
POLY-A=Yes.

FEATURES
source Location/Qualifiers
1..544
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="UI-1-BC1P-ay1-f-03-0-UI"
/clone_lib="NCI-CGAP_P13"
/tissue_type="Placenta"
/dev_stage="8-9 weeks"
/lab_host="DH10B (Life Technologies)"
/note="Organ: Placenta; Vector: pTY73-Pac (Pharmacia) with
a modified polylinker; Site_1: Ecor I; Site_2: Not I;
NCI-CGAP_P13 is a subcloned cDNA library constructed
according to Bonaldo, Lennon and Soares, Genome Research,
6:791-806, 1996. First strand cDNA synthesis was primed
with an oligo-dT primer containing a Not I site. Double

stranded cDNA was ligated to an Ecor I adaptor, digested
with Not I, and cloned directionally into pTY73-Pac
vector. The oligonucleotide used to prime the synthesis of
first-strand cDNA contains a library tag sequence that is
located between the Not I site and the (dT)18 tail. The
sequence tags for this library are GA, AGGA. For
additional information, contact: Bento Soares,
bento-soares@uiowa.edu
TAG_LIB=UI-1-BC1P
TAG_TISSUE=Placenta human 8 week
TAG_SEQ=GA*

BASE COUNT 163 a 106 c 101 g 174 t
ORIGIN
Query Match 13.9%; Score 537.8; DB 14; Length 544;
Best Local Similarity 99.6%; Pred. No. 4.7e-103;
Matches 539; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy 543 TGTGGAAGAGTGTGAGAAATCTTTTGTGAGGAACCTGATCTGCGCAAGTTCCA 602
Db 544 TGTGGAAGAGTGTGAGAAATCTTTTGTGAGGAACCTGATCTGCGCAAGTTCCA 485
Qy 603 TGTAGCACTAATCATTTTATAGACAGCAACCACTGAAAGTTGTAAGATCATGTTTC 662
Db 484 TGTAGCACTAATCATTTTATAGACAGCAACCACTGAAAGTTGTAAGATCATGTTTC 425
Qy 663 TATGACACCTACCATGCTGCTATACAGACTGCTATACAGCACTTTTAAATGATGTCT 722
Db 424 TATGACACCTACCATGCTGCTATACAGACTGCTATACAGCACTTTTAAATGATGTCT 365
Qy 723 AAAGAACTAAATGCAATACCCATGCTTGAAGTGAAGATTTATCTTTAGAAATGTC 782
Db 364 AAAGAACTAAATGCAATACCCATGCTTGAAGTGAAGATTTATCTTTAGAAATGTC 305
Qy 783 TATTGGAACCACTTTTGACAGCAACATCCGCCATTTATCTGATCTTTGGACACGCT 842
Db 304 TATTGGAACCACTTTTGACAGCAACATCCGCCATTTATCTGATCTTTGGACACGCT 245
Qy 843 TGGAGCCAGAGCTAATCCCTTATGATGAGATTTGAGAACTTTGCTGCGAGTA 902
Db 244 TGGAGCCAGAGCTAATCCCTTATGATGAGATTTGAGAACTTTGCTGCGAGTA 185
Qy 903 TCTCTCTCACTATGATTTGTGTCACATTTTCTTAACTTTTGGAAATCTGTGAGACACGA 962
Db 184 TCTCTCTCACTATGATTTGTGTCACATTTTCTTAACTTTTGGAAATCTGTGAGACACGA 125
Qy 963 AAAAGCTTTTGTGACAGATTCAGTTGGCTGCTTCTTGAATCCAGCACTCGATGTTAT 1022
Db 124 AAAAGCTTTTGTGACAGATTCAGTTGGCTGCTTCTTGAATCCAGCACTCGATGTTAT 65
Qy 1023 AAATGCTCGAGCAAGGTTTATCATCTTCCAGATGCCAAATGATTAAGAAAGAAAAAT 1082
Db 64 AAATGCTCGAGCAAGGTTTATCATCTTCCAGATGCCAAATGATTAAGAAAGAAAAAT 5
Qy 1083 A 1083
Db 4 A 4

RESULT 7
AM977575/c 581 bp mRNA linear EST 02-JUN-2000
LOCUS AM977575
DEFINITION EST389684 MAGE resequences, MAGE Homo sapiens cDNA, mRNA sequence.
ACCESSION AM977575
VERSION AM977575.1 GI:8168828
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 581)
Hegde, P., Qi, R., Abernathy, K., Dharap, S., Gaspard, R., Gay, C., Holt,
I.E., Saeed, A.I., Sharov, V., Lee, N.H., Yeatman, T.J. and

Db 266 GCTCTTGGTGGTCCAGGTCACAGTACTGATTTGTGCAAGTATGACCGAATGTTCCAG 207

Qy 1270 CTGAGACATATATACACTCTTGGACGAGCGCTTCTTATTGAGGCTCTACAGAAACC 1329

Db 206 CTGAGACATATATACACTCTTGGACGAGCGCTTCTTATTGAGGCTCTACAGAAACC 147

Qy 1330 TTGGAGAGGA--TGACAAAGCTGAAAGAGTGTGA---TGAAATTTTGAAGAGAGC 1383

Db 146 TTGGAGAGGATACCAAGAGCTGAAAGAGTGTGAAGATGAATTTTGAAGAGAGAGC 87

Qy 1384 ACTGCAAA-GAGAAATTAG-GAAATCTCACAAAGACCTTAAGACCCCAAAACAAAGAGC 1441

Db 86 ACTGCAAAAGAGATTTAGGAATCTCACAAAGACCTTAAGACCCCAAAACAAAGAGC 27

Qy 1442 GGGCTTTCTACCAAGAAAGACCCCTC 1467

Db 26 GGGCTTTCTACCAAGAAAGACCCCTC 1

RESULT 9

LOCUS B0011470 530 bp mRNA linear EST 26-MAR-2002

DEFINITION UI-1-BC1P-ase-e-04-0-UI.s1 NCI_CGAP_P13 Homo sapiens cDNA clone

ACCESSION B0011470

VERSION B0011470

KEYWORDS UI-1-BC1P-ase-e-04-0-UI 3', mRNA sequence.

SOURCE EST.

ORGANISM B0011470.1 GI:19736371

human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 530)

NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

National Cancer Institute, Cancer Genome Anatomy Project (CGAP).

Tumor Gene Index

Unpublished (1997)

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-r@mail.nih.gov

Tissue Procurement: Dr. Steven Brown

cDNA library preparation: Dr. M. Bento Soares, University of Iowa

DNA library Arrayed by: Dr. M. Bento Soares, University of Iowa

DNA Sequencing by: Dr. M. Bento Soares, University of Iowa

Clone Distribution: Clone distribution information can be obtained from Dr. M. Bento Soares, bento-soares@uiowa.edu

Seq primer: M13 FORWARD

PolyA=yes.

Location/Qualifiers

1. 530

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="UI-1-BC1P-ase-e-04-0-UI"

/clone_1lb="NCI_CGAP_P13"

/tissue_type="Placenta"

/dev_stage="8-9 weeks"

/lab_host="PH10B (Life Technologies)"

/note="Organ: Placenta; Vector: pRT73-Pac (Pharmacia) with a modified polylinker; Site_1: EcoR I; Site_2: Not I; NCI-CGAP_P13 is a subtracted cDNA library constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pRT73-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tags for this library are GA, AGCA. For additional information, contact: Bento Soares, bento-soares@uiowa.edu

TAG_LIB=UI-1-BC1P

TAG_TISSUE=Placenta human 8 week

TAG_SEQ=GA"

BASE COUNT 160 a 99 c 101 g 170 t

ORIGIN

Query Match 13.5%; Score 520.6; DB 14; Length 530;

Best Local Similarity 99.2%; Pred. No. 2e-99;

Matches 523; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 557 TGAGAAATCTTTTGTGAGAAAGTATATCTGTGCGCAGGTTCCATGACGATTAAC 616

Db 530 TGAGAAATCTTTTGTGAGAAAGTATATCTGTGCGCAGGTTCCATGACGATTAAC 471

Qy 617 CATTTTACAGACAGCAAAACCTGAAGTTGTAAGAAATCCATGTTCTATGACACTACCA 676

Db 470 CATTTTACAGACAGCAAAACCTGAAGTTGTAAGAAATCCATGTTCTATGACACTACCA 411

Qy 677 TGGTGGCTATACAGACTGATACAGATTTTAAATGATGCTTAAGAACTAAAT 736

Db 410 TGGTGGCTATACAGACTGATACAGATTTTAAATGATGCTTAAGAACTAAAT 351

Qy 737 GCCATTAACCCATCGCTTGAAGTGAAGATTTATCTTTAGAAATGCTATTGAAAAACCTT 796

Db 350 GCCATTAACCCATCGCTTGAAGTGAAGATTTATCTTTAGAAATGCTATTGAAAAACCTT 291

Qy 797 TTGACAAAGCAATCCGCCATTTATCTGATCTTTTGGCAGCAGCTTGGAGCCAGACTA 856

Db 290 TTGACAAAGCAATCCGCCATTTATCTGATCTTTTGGCAGCAGCTTGGAGCCAGACTA 231

Qy 857 AATCCTTAGTTCAGATTTGAAGATTTAGCAACTTGTCTGAGATCTCTCATGATG 916

Db 230 AATCCTTAGTTCAGATTTGAAGATTTAGCAACTTGTCTGAGATCTCTCATGATG 171

Qy 917 ATTGTGTCAATTTCTTAATCTCTGGAATCTGTGAGCAACGCAAAAGCTTTGGTTC 976

Db 170 ATTGTGTCAATTTCTTAATCTCTGGAATCTGTGAGCAACGCAAAAGCTTTGGTTC 111

Qy 977 AGAATTCAGTTGGCTGTTTCTTGACTCCAGACCTCGATGTTTAAATGCTCGAGCA 1036

Db 110 AGAATTCAGTTGGCTGTTTCTTGACTCCAGACCTCGATGTTTAAATGCTCGAGCA 51

Qy 1037 GGGTTATCATCTCCAGATGCCAAATGATTAAGAAAAATA 1083

Db 50 GGGTTATCATCTCCAGATGCCAAATGATTAAGAAAAATA 4

RESULT 10

LOCUS B1555512 829 bp mRNA linear EST 05-SEP-2001

DEFINITION 603239318F1 NCI_CGAP_Mam3 Mus musculus cDNA clone IMAGE:5292221 5', mRNA sequence.

ACCESSION B1555512

VERSION B1555512

KEYWORDS B1555512.1 GI:15442826

SOURCE EST.

ORGANISM B1555512

house mouse.

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 829)

NIH-MGC <http://mgc.nci.nih.gov/>.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-r@mail.nih.gov

Tissue Procurement: Lothar Hennighausen Ph.D., Chu-Xia Deng Ph.D.

cDNA library preparation: Life Technologies, Inc.

DNA library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: MGC Genomics, Inc.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>

Plate: L1AM11739 row: C column: 06

High quality sequence start: 2

High quality sequence stop: 753.

Location/Qualifiers

1. 829

source

```

/organism="Mus musculus"
/strain="129,C57BL/6J,FVB/N"
/db.xref="taxon:10090"
/clone="IMAGE:5292221"
/tissue_type="tumor, gross tissue"
/dev_stage="10 months"
/lab_host="DH10B"
/note="Organ: mammary; Vector: pCMV-Sport6; Site: 1; SalI;
Site: 2; NotI; Cloned unidirectionally. Primer: Oligo dT.
Library constructed by Life Technologies. Investigators
providing samples: Lothar Hennighausen/Chu-Xia Deng, NIH
Reference for transgenic model: Xu et al., Nature Genetics
22, 37-43 (1999)."
BASE COUNT      209 a      214 c      216 g      190 t
ORIGIN
Query Match      13.3%; Score 513.6; DB 13; Length 829;
Best Local Similarity 83.1%; Pred. No. 6e-98;
Matches 680; Conservative 0; Mismatches 129; Indels 9; Gaps 8;
OY 5 GAGCTTCATGGAGTACGAGCGCGGCTCG-ACGGATTGCCATGGCCCGCTGCTG-GA 62
Db 1 GGGGCCCCATGGAGCCAGCCCTGTGTGGCGCAACGGCGAGTATGCGCCGCTGTGTGA 60
OY 63 GTACGAG-CGACAGCTGGTGTGTAAGTCTGACACTGACGAGCGGTAGTGTGCGCCC 121
Db 61 GTACGAGATGGCAGCAGAGGTGCTGGAACCTGCTGACAGCAGCGCGGTGTGTGCCC 120
OY 122 GCGGCGCTGGCGCGGAGC-GGCTCCTGTACCACTTTCTCCAGCTGCACGCCAGCC 180
Db 121 GCGGACTGGGACCCAGCTGGCTCTTACCACTTTCTCCAGCTGCACGCCAGCC 180
OY 181 TGCCGATGCTGGTGTCTCAACGACGCGCGCGAGGAGGATTTATCAATGAGCTG 240
Db 181 TGCCGATGCTGGTGTCTCAACGACGCGCGCGAGGAGGATTTATCAATGAGCTG 240
OY 241 AAGATAGAGAGAGTTGAACACCTCCCTCCGCTGTACCAATGAATCACAACAAGT 300
Db 241 AAGATAGAGAGAGTTGAACATCTCCCTCGACAGTGAACAATGAATGCAAGTAACGT 300
OY 301 CGCTGTGAAGTTTACACACAAGGTGTGTATTTTGGACAGTAGATTTGTGTGTT 360
Db 301 CGCTGTGAAGTTTACACACAGGAGGTGTATTTATTTGGACACGCAATCTCTGTGTT 360
OY 361 GACTTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 420
Db 361 GATTTCCTGACAGTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATG 420
OY 421 CACAGATTAATGCA-GTCTTGTCAAGAAGCATTCCTTGGCCCTTTTCCGACAAAAA 479
Db 421 CACAGATTAATGCAAGTCTCTGCGAGAGACCTTCATCTGCGCCCTTCCGACAAAAA 480
OY 480 CAACGCTGTTTATTAAGCTTTACAGACATGCTGTG-CCCTTGATGATGTTT 538
Db 481 CAAGGCGGGTTCATCAAGCTTTCACGACACGCTGTGCGCTTTGACACTGCTTTT 540
OY 539 GTCATGTGGAAGAGTGTATGAGAAATCTTTTGTGAGAAACTGTATCTGTGCGCAAGT 598
Db 541 GTCAAGTGTGAAGAGTGTATGCGGAACCTTTTGTGAGAGAGTCTACTGTGCGCAAGT 600
OY 599 TCCATGTAGCAAGTAACTCAATTTTGAAGACAGCAAACTGGAAGTTTGAAGATTCATG 658
Db 601 TCCATGTAGCAAGTAACTCAATTTTGAAGACAAACAGCTGGAAGTTTGAAGATTCACG 660
OY 659 TTTTCATGACACTTACCATGCTGTGCTATACAG-ACCTGATATACAGAC-ATTTAATG 715
Db 661 TGTTCATGACTCCCGCCAGCTGTGCTATGAGAGGCGCATTAAGACATCTTATACATG 720
OY 716 CATGCTTAAGAGAACTAAATGCCATAACCATGCTGTAAGTGAAGATTTATCTTAAAG 775
Db 721 CATGCTTAAGAGAACTAAATGTACACACCGCTGACTTGAAGTGAAGATTTATCTTAAAG 780

```

```

OY 776 AAAATGCTATTGAAAACCTTTTGACAGACATCCGC 813
Db 781 AACAGCTCTTGGGAAGCATTCGACAGACATCCGC 818

```

```

RESULT 11
LOCUS      BQ310815          515 bp      mRNA      linear      EST 16-MAY-2002
DEFINITION OY-BN0046-150300-121-b11 BN0046 Homo sapiens cDNA, mRNA sequence.
ACCESSION  BQ310815
VERSION     BQ310815.1 GI:20866223
KEYWORDS
SOURCE      human.
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE  1 (bases 1 to 515)
AUTHORS    Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bouidin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H.,
Brunstein,A., deoliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
TITLE       Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
JOURNAL     Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE     20202663
COMMENT     Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 0:539-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/ICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?rl-QV3&t2-QV3-BN0046-
150300-121-b11&t3=2000-03-15&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 17
High quality sequence stop: 515.
Location/Qualifiers
1..515
/organism="Homo sapiens"
/db.xref="taxon:9606"
/clone_lib="BN0046"
/dev_stage="Adult"
/note="Organ: breast-normal; Vector: puc18; Site:1; SmaI;
Site:2; SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the puc 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."

```

```

BASE COUNT      131 a      134 c      118 g      132 t
ORIGIN
Query Match      13.0%; Score 499.6; DB 14; Length 515;
Best Local Similarity 98.2%; Pred. No. 5.4e-95;
Matches 505; Conservative 0; Mismatches 9; Indels 0; Gaps 0;
OY 1944 AGAAGAGACAGCAGATGAAACAACCTTAGCTAGTAAAGGCGACAGCATCTGCGA 2003
Db 2 AGAAAGCAGAAAGCAGATGAAACAACCTTAGCTAGTAAAGGCGACAGCATCTGCGA 61
OY 2004 TGTTCACACTGACACTCGGAAAGCGGTGGCCAGAGACAGATGCTACACAGCAAGAT 2063
Db 62 TGTTCACACTGACACTCGGAAAGCGGTGGCCAGAGACAGATGCTACACAGCAAGAT 121
OY 2064 AGTGTGATATGCGTGAATTTGCAAGAGAGTTCATCTGATCCATTCGCGGGGAT 2123
Db 122 AGTGTGATATGCGTGAATTTGCAAGAGTTCATCTGATCCATTCGCGGGGAT 181

```


QY 2124 TGACATTGAACCCGTCAGTTAGAGGTGGAGATTACATCCCTCAGTCACAGAAATGCGCT 2183
 |||||
 DB 182 TGACATTGAACCCGTCAGTTAGAGGTGGAGATTACATCCCTCAGTCACAGAAATGCGCT 241
 |||||
 QY 2184 GGAGCCCAAGAGATGATGATTAATCGGCTTTAAATACGGCCGCTCTACAGCCA 2243
 |||||
 DB 242 GGAGCCCAAGAGATGATGATTAATCGGCTTTAAATACGGCCGCTCTACAGCCA 301
 |||||
 QY 2244 GTGCATCTCCATCTCCCGCTACATCAAGAGCTCCCGCTCTCTATGATGATGACCTAG 2303
 |||||
 DB 302 GTGCATCTCCATCTCCCGCTACATCAAGAGCTCCCGCTCTCTATGATGATGACCTAG 361
 |||||
 QY 2304 CAAGCCCTTCTCTCTACCTCCCGAGTGCCTTTTACAGAGATCTCCAGCATGACAT 2363
 |||||
 DB 362 CAAGCCCTTCTCTCTACCTCCCGAGTGCCTTTTACAGAGATCTCCAGCATGACAT 421
 |||||
 QY 2364 TAGTCCCAACATCAGTCTTCTTACCTTCACTTCCCAAGATGATGATGATGATGATG 2423
 |||||
 DB 422 TAGTCCCAACATCAGTCTTCTTACCTTCACTTCCCAAGATGATGATGATGATGATG 481
 |||||
 QY 2424 CTCTCTCTACATGCAAGCGGCGAGTGTGTTGAGAG 2457
 |||||
 DB 482 CTCTCTCTACATGCAAGCGGCGAGTGTGTTGAGAG 515
 |||||

RESULT 12
 BM671280/c 524 bp mRNA 1linear EST 27-FEB-2002
 LOCUS
 DEFINITION
 UI-E-CQ1-agg-h-17-0-UI-51 UI-E-CQ1 Homo sapiens cDNA clone
 ACCESSION
 BM671280
 VERSION
 BM671280.1 GI:18981178
 KEYWORDS
 EST.
 SOURCE
 human.
 ORGANISM
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 524)
 REFERENCE
 Ronaldo, M.F., Lennon, G. and Soares, M.B.
 Normalization and subtraction: two approaches to facilitate gene
 TITLE
 discovery
 JOURNAL
 Genome Res. 6 (9), 791-806 (1996)
 MEDLINE
 97044477
 COMMENT
 Contact: Soares, MB
 Program for Rat Gene Discovery and Mapping
 University of Iowa
 451 Eckstein Medical Research Building Iowa City, IA 52242, USA
 Tel: 319 335 8250
 Fax: 319 335 9565
 Email: msoares@blue.weeg.uiowa.edu
 Tissue Procurement: Dr. Gregg Hageman
 cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: Researchers may obtain clones from Research
 Genetics (www.resgen.com).
 Seq primer: M13 forward
 POLY-A-tail.

FEATURES
 source
 Location/Qualifiers
 1..524
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="UI-E-CQ1-agg-h-17-0-UI"
 /clone_lib="UI-E-CQ1"
 /tissue_type="Optic nerve"
 /dev_stage="adult"
 /lab_host="DH10B (Life Technologies) (T1 phage resistant)"
 /note="Organ: eye; Vector: pTZ19-Pac (pharmacia) with a
 modified polylinker; Site_1: EcoR I; Site_2: Not I;
 UI-E-CQ1 is a normalized cDNA library containing the
 following tissue(s): optic nerve. The library was
 constructed according to Ronaldo, Lennon and Soares,

Genome Research, 6:791-806, 1996. First strand cDNA
 synthesis was primed with an oligo-dT primer containing a
 Not I site. Double stranded cDNA was ligated to an EcoR I
 adaptor, digested with Not I, and cloned directionally
 into pTZ19-Pac vector. The oligonucleotide used to prime
 the synthesis of first-strand cDNA contains a library tag
 sequence that is located between the Not I site and the
 (GT)₁₈ tail. The sequence tag for this library is
 CCATTAAGTC. This library was created for the program, Gene
 Discovery in the Visual System, supported by National Eye
 Institute (NEI).
 TAG_LIB-UI-E-CQ1
 TAG_TISSUE-human optic nerve
 TAG_SEQ-CCATTAAGTC

BASE COUNT 119 a 136 c 127 g 142 t

ORIGIN

Query Match 12.9%; Score 499; DB 13; Length 524;
 Best Local Similarity 100.0%; Pred. No. 7.2e-95;
 Matches 499; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 14 TGGAGTCAGGGGACGCGGCTCGACGATTCGATGGCGGCTGCTGGAGTACGAGCGAC 73
 |||||
 DB 515 TGGAGTCAGGGGACGCGGCTCGACGATTCGATGGCGGCTGCTGGAGTACGAGCGAC 456
 |||||
 QY 74 AGCTGGTGTGGAAGTCTGACACATGACGCGGCTAGTGTGCGCCGCGGCTCGGCG 133
 |||||
 DB 455 AGCTGGTGTGGAAGTCTGACACATGACGCGGCTAGTGTGCGCCGCGGCTCGGCG 396
 |||||
 QY 134 CGGACCGGCTCTCTACACATTTCTCCAGCTGCACTGCCACGAGCCCTGCTGCTGG 193
 |||||
 DB 395 CGGACCGGCTCTCTACACATTTCTCCAGCTGCACTGCCACGAGCCCTGCTGCTGG 336
 |||||
 QY 194 TGTCTCAACAGCAGCGCGCGGAGAGAGATTTATCATATCAGCTTAATAGAGAG 253
 |||||
 DB 335 TGTCTCAACAGCAGCGCGCGGAGAGAGATTTATCATATCAGCTTAATAGAGAG 276
 |||||
 QY 254 TTGAACACCTCCCTCGCGGCTGTAAACAATGAATACACACAGCAGCTATGAAGTT 313
 |||||
 DB 275 TTGAACACCTCCCTCGCGGCTGTAAACAATGAATACACACAGCAGCTATGAAGTT 216
 |||||
 QY 314 ACACACAAGGTGTGTATTTTGCAGCAAGTAGAGATCTTGTGCTTCTTGACTG 373
 |||||
 DB 215 ACACACAAGGTGTGTATTTTGCAGCAAGTAGAGATCTTGTGCTTCTTGACTG 156
 |||||
 QY 374 ATGAAATACCTTCAGATTTTAACTGCGATCTTGTGCTTATAGAGCCCAAGATATCG 433
 |||||
 DB 155 ATGAAATACCTTCAGATTTTAACTGCGATCTTGTGCTTATAGAGCCCAAGATATCG 96
 |||||
 QY 434 AGCTTGTCAAGAGCATTCATCTTGGCGCTCTTGCAGAAACCAAGCGTGTGTTA 493
 |||||
 DB 95 AGCTTGTCAAGAGCATTCATCTTGGCGCTCTTGCAGAAACCAAGCGTGTGTTA 36
 |||||
 QY 494 TTAAGCTTTACAGACAA 512
 |||||
 DB 35 TTAAGCTTTACAGACAA 17
 |||||

RESULT 13
 BE818393/c 618 bp mRNA 1linear EST 21-SEP-2000
 LOCUS
 BE818393
 DEFINITION
 RCS-BN0282-010600-031-B08 BN0282 Homo sapiens cDNA, mRNA sequence.
 ACCESSION
 BE818393
 VERSION
 BE818393.1 GI:10250627
 KEYWORDS
 EST.
 SOURCE
 human.
 ORGANISM
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 618)
 REFERENCE
 Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,
 Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,
 Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H.,

TITLE
Brunstein, A., deOliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and Simpson, A.J.
Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663

JOURNAL MEDLINE
Contact: Simpson, A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br

COMMENT
This sequence was derived from the FAPESP/LICR Human Genome Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?l=6t2-RC5-BN0282-010600-031-B08&t3=2000-06-01&t4=1)
Seq primer: puc 18 forward
High quality sequence stop: 427.

FEATURES

SOURCE
1. 618
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="BN0282"
/dev_stage="Adult"
/note="Organ: breast normal; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from ORSTS PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."
Location/Qualifiers
132 a 155 c 114 g 217 t

BASE COUNT
ORIGIN

Query Match 12.6%; Score 485.6; DB 12; Length 618;
Best Local Similarity 99.0%; Pred. No. 4.9e-92;
Matches 499; Conservative 0; Mismatches 4; Indels 1; Gaps 1;

QY 1221 TCCAGGTCACATGATGATTTGTCAGAGATGACCGAATGTTCCAGCTGAGAGACTA 1280
DB 512 TTCAGGTCACATGATGATTTGTCAGAGATGACCGAATGTTCCAGCTGAGAGACTA 453
QY 1281 TATCACTCTTGAGCGCA-GGCTTCTTATGAGGCTCTACAGAAACCTTTGAGAGG 1339
DB 452 TATCACTCTTGAGCGCGGCTTCTTATGAGGCTCTACAGAAACCTTTGAGAGG 393
QY 1340 ATAGCAAGCTGAGAAAGTCTGATGAATTTAGGAAGAGACAGTTCAAGAGAAATTA 1399
DB 392 ATAGCAAGCTGAGAAAGTCTGATGAATTTAGGAAGAGACAGTTCAAGAGAAATTA 333
QY 1400 GGAATCTCACAAGAAAGCTTAAGACCCCAAAACAAAGAGCGGCTTCTACCAAGAAA 1459
DB 332 GGAATCTCACAAGAAAGCTTAAGACCCCAAAACAAAGAGCGGCTTCTACCAAGAAA 273
QY 1460 GAACCTCAAAAAAGAAAGAGTGTGACCTTAACCTCAATGGTGAAGAAACCTGTAAG 1519
DB 272 GAACCTCAAAAAAGAAAGAGTGTGACCTTAACCTCAATGGTGAAGAAACCTGTAAG 213
QY 1520 AACTGGAAGAGAGAGATGTCGAGAGAGATATCGTGAAGAAATTAAGCAGTACCCAG 1579
DB 212 AACTGGAAGAGAGAGATGTCGAGAGAGATATCGTGAAGAAATTAAGCAGTACCCAG 153
QY 1580 AAAGCTCCCGGGAAGAAATTAAGCATGAAGATTTGATTAATTTGTCATCGGATGCTG 1639
DB 152 AAAGCTCCCGGGAAGAAATTAAGCATGAAGATTTGATTAATTTGTCATCGGATGCTG 93
QY 1640 CTTTCGGAATCTGGAAGAAACCCCTCACTATCATCGCTCTGCTGGTTCAGCGCAGC 1699
DB 92 CTTTCGGAATCTGGAAGAAACCCCTCACTATCATCGCTCTGCTGGTTCAGCGCAGC 33

QY 1700 CCTATGCTCTGACAGGGTACTAC 1723
DB 32 CCTATGCTCTGACAGGGTACTAC 9

RESULT 14
B0944411
LOCUS
DEFINITION
IMAGE:6440053 5', mRNA sequence.
ACCESSION
B0944411
VERSION
B0944411.1 GI:22359889
KEYWORDS
EST.
SOURCE
house mouse.
ORGANISM
Mus musculus

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgap@femail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Agencourt Bioscience Corporation
clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
Plate: LLM13961 row: a column: 14
High quality sequence stop: 536.

FEATURES

SOURCE
1. 923
/organism="Mus musculus"
/strain="FVB/N-3"
/db_xref="taxon:10090"
/clone="IMAGE:6440053"
/clone_lib="NCI CGAP Mam2"
/tissue_type="tumor; biopsy sample"
/dev_stage="5 months"
/lab_host="DH10B"
/note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI; Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT. Library constructed by Life Technologies. Investigator providing samples: Gilbert Smith, NIH"
Location/Qualifiers
211 a 281 c 266 g 164 t
others

BASE COUNT
ORIGIN

Query Match 12.5%; Score 482.2; DB 14; Length 923;
Best Local Similarity 78.7%; Pred. No. 2.6e-91;
Matches 588; Conservative 0; Mismatches 18; Indels 1; Gaps 1;

QY 1823 GGGTTTACTTCTTATATACGAGGTTCACTGAGAAACAACCTATCTCACTGCTTTC 1882
DB 17 GGGTTTACTTCTTATATACGAGGTTCACTGAGAAACAACCTATCTCACTGCTTTC 76
QY 1883 GGAAGAAGAAAGAGCTTTGAAAAACATCAAGGAAGAAAGCAAGCATGGTTCCTG 1942
DB 77 GGAAGAAGAAAGAGCTTTGAAAAACATCAAGGAAGAAAGCAAGCATGGTTCCTG 136
QY 1943 AAGAAGAGAGAGAGATGAACAACCTAGACCTAGTAAAGGCACAGCATCTGCAG 2002
DB 137 AAGAAGAGAGAGAGATGAACAACCTAGACCTAGTAAAGGCACAGCATCTGCAG 196
QY 2003 ATGTTTCCACTGACCTCGGAAGCCGCTGGCCAGGAACAGCAATGTACACAGCAACA 2062
DB 197 ATGTTTCCACTGACCTCGGAAGCCGCTGGCCAGGAACAGCAATGTACACAGCAACA 256
QY 2063 TAGTTTGGAGATTCGGGAATTTGGAAGTGAAGCTTCATCTCTATCATCTGCTGGGCA 2122
DB 257 TAGTTTGGAGATTCGGGAATTTGGAAGTGAAGCTTCATCTCTATCATCTGCTGGGCA 316

1	2718	70.5	2718	4	US-09-651-656-14	Sequence 14, Appl
2	2718	70.5	2718	4	US-09-650-855-14	Sequence 14, Appl
3	1101	28.5	1101	4	US-09-651-656-12	Sequence 12, Appl
4	1101	28.5	1101	4	US-09-650-855-12	Sequence 12, Appl
5	78	2.0	7218	1	US-08-232-643-14	Sequence 14, Appl
6	68.6	1.8	80246	4	US-09-078-294-4	Sequence 4, Appl
7	68.2	1.8	162450	4	US-09-345-882-1	Sequence 1, Appl
8	68	1.8	45716	4	US-08-965-048-5	Sequence 5, Appl
9	68	1.8	45969	4	US-08-965-048-6	Sequence 6, Appl
10	67.8	1.8	168575	4	US-09-426-290-1	Sequence 1, Appl
11	67.2	1.7	80595	4	US-09-078-294-3	Sequence 3, Appl
12	67	1.7	176373	3	US-09-128-815-17	Sequence 17, Appl
13	66.6	1.7	84455	4	US-09-797-906-3	Sequence 3, Appl
14	66.4	1.7	45716	4	US-08-965-048-5	Sequence 5, Appl
15	66.4	1.7	45969	4	US-08-965-048-6	Sequence 6, Appl
16	66.2	1.7	14796	4	US-08-975-080-35	Sequence 35, Appl
17	66.2	1.7	14796	4	US-09-630-706-10	Sequence 10, Appl
18	66.2	1.7	14796	4	US-09-630-706-10	Sequence 10, Appl
19	66.2	1.7	55827	4	US-09-813-133A-3	Sequence 3, Appl
20	66.2	1.7	70000	4	US-09-351-896-3	Sequence 3, Appl
21	66	1.7	6792	4	US-09-374-454-20	Sequence 20, Appl
22	66	1.7	8224	2	US-09-010-398-14	Sequence 14, Appl
23	66	1.7	8224	4	US-09-366-260-14	Sequence 14, Appl
24	66	1.7	246240	2	US-08-724-394A-20	Sequence 20, Appl
25	66	1.7	246240	2	US-08-724-394A-20	Sequence 21, Appl
26	66	1.7	246240	2	US-08-724-394A-22	Sequence 22, Appl
27	65.8	1.7	35100	1	US-08-306-691B-19	Sequence 19, Appl

C	28	65.8	1.7	35100	5	PCT-US93-06251-15	Sequence 19, Appl
C	29	65.8	1.7	40352	3	US-08-846-1110-15	Sequence 15, Appl
C	30	65.8	1.7	40352	4	US-09-443-077-15	Sequence 15, Appl
C	31	65.8	1.7	50000	4	US-09-146-053-4	Sequence 4, Appl
C	32	65.6	1.7	79629	4	US-09-729-995-3	Sequence 3, Appl
C	33	65.6	1.7	43755	3	US-08-742-185-101	Sequence 101, Appl
C	34	65.4	1.7	99300	4	US-09-798-096-10	Sequence 10, Appl
C	35	65	1.7	11811	4	US-09-078-294-7	Sequence 7, Appl
C	36	64.8	1.7	2907	2	US-09-018-628-17	Sequence 17, Appl
C	37	64.8	1.7	2907	3	US-09-373-378-17	Sequence 17, Appl
C	38	64.8	1.7	2907	4	US-09-018-635-26	Sequence 26, Appl
C	39	64.8	1.7	2907	4	US-09-467-654-3	Sequence 3, Appl
C	40	64.8	1.7	16389	4	US-09-741-154-3	Sequence 3, Appl
C	41	64.5	1.7	1001	4	US-09-641-638-354	Sequence 354, Appl
C	42	64.6	1.7	2908	3	US-08-487-799-1	Sequence 1, Appl
C	43	64.6	1.7	14581	4	US-08-520-3730-4	Sequence 4, Appl
C	44	64.6	1.7	14636	4	US-09-173-914-6	Sequence 6, Appl
C	45	64.4	1.7	3366	4	US-09-343-650-2	Sequence 2, Appl

ALIGNMENTS

```

? RESULT 1
? US-09-651-656-14
? Sequence 14 Application US/09651656
? Patent No. 6340566
? GENERAL INFORMATION:
? APPLICANT: MCCUTHEN-MALONEY, SANDRA
? APPLICANT: LAWRENCE LIVERMORE NATIONAL LABORATORY
? TITLE OF INVENTION: DETECTION AND QUANTITATION OF SINGLE NUCLEOTIDE
? TITLE OF INVENTION: POLYMORPHISMS, DNA SEQUENCE VARIATIONS, DNA MUTATIONS
? TITLE OF INVENTION: DNA DAMAGE AND DNA MISMATCHES
? FILE REFERENCE: IL-10689
? CURRENT APPLICATION NUMBER: US/09/651,656
? CURRENT FILING DATE: 2000-08-29
? PRIOR APPLICATION NUMBER: 60/192,764
? PRIOR FILING DATE: 2000-03-28
? NUMBER OF SEQ. ID NOS: 106
? SOFTWARE: PatentIn Ver. 2.1
? SEQ. ID NO 14
? LENGTH: 2718
? TYPE: DNA
? ORGANISM: Homo sapiens
? US-09-651-656-14

```

Query Match	70.5%;	Score 2718;	DB 4;	length 2718;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 2718; Conservative	0;	Mismatches	0;	Indels 0; Gaps 0;

Y	40	ATGCGCCGCGCTGCTGAGAGTACGAGCGACAGTGGTGTGGAACAGCTCGACACTGACGAGGG	107
Db	1	ATGCGCGCGGCTGCTGAGAGTACGAGCGACACTGGTGTGGAACAGCTCGACACTGACGAGGG	60
QY	106	CTATAGTGTGCGCGCGCGGCGCTGGGCGGACCGGCGCTGACACAGTTTCTTCACACTG	165
Db	61	CTATAGTGTGCGCGCGCGGCGCTGGGCGGACCGGCGCTCTTCACACTTTCCTCAGCTG	120
QY	166	CACGCGCACCCAGCGCTGCTGGTGCTGTCTCAACACGACGCGCGCGGAGGAGTAT	225
Db	121	CACGCGCACCCAGCGCTGCTGGTGCTGTCTGTCTCAACACGACGCGCGCGGAGGAGTAT	180
QY	226	TTTATTCATATAGCTGGAAGATATGAAGGAGTTGACACCTCCCTCGCGCTGTACACAAATGAA	285
Db	181	TTTATTCATATAGCTGGAAGATATGAAGGAGTTGACACCTCCCTCGCGCTGTACACAAATGAA	240
QY	286	ATCAACAAGCAACAGTCCGCTATGAAGTTTAAACAACAAAGTGCTGTATTTATTCGACAAGT	345
Db	241	ATCAACAAGCAACAGTCCGCTATGAAGTTTAAACAACAAAGTGCTGTATTTATTCGACAAAGT	300
QY	346	AGGATACCTTGTGGTGTGACTCTTGACATGATGAATACCTTCAGATTTAAATTACTGCGATC	405
Db	301	AGGATACCTTGTGGTGTGACTCTTGACATGATGAATACCTTCAGATTTAAATTACTGCGATC	360

QY 406 TTGGTGTATAGAGCCACAGATATATGAGTCTTGTCAAGAGCATTCATCTTGCGCTC 465
| | | | |
Db 361 TTGGTGTATAGAGCCACAGATATATGAGTCTTGTCAAGAGCATTCATCTTGCGCTC 420
QY 466 TTTCGCCAGAAAAACAACGTGGTTTATTAAGCTTTCACAGACAATGCTGTGCTTT 525
| | | | |
Db 421 TTTCGCCAGAAAAACAACGTGGTTTATTAAGCTTTCACAGACAATGCTGTGCTTT 480
QY 526 GATACGTGTTTGTCTATGTGGAAAAAGTATGAAATCTTTTGTGAGAAACTGTAT 585
| | | | |
Db 481 GAATACGTGTTTGTCTATGTGGAAAAAGTATGAAATCTTTTGTGAGAAACTGTAT 540
QY 586 CTGTGGCCAGAGTTCCATGTAGCAATTAACCTATTTTTCAGACAGCAAAACCTGAAGTT 645
| | | | |
Db 541 CTGTGGCCAGAGTTCCATGTAGCAATTAACCTATTTTTCAGACAGCAAAACCTGAAGTT 600
QY 646 GTAGAAATCCATGTTTCTATGACACCTACCATGCTTGTCTATAGACACTGCTATACAGT 705
| | | | |
Db 601 GTAGAAATCCATGTTTCTATGACACCTACCATGCTTGTCTATAGACACTGCTATACAGT 660
QY 706 ATTTTAAATGCAATGCTTAAAGAACTAAATGCCATTAACCATGCTTGAAGTGAAGAT 765
| | | | |
Db 661 ATTTTAAATGCAATGCTTAAAGAACTAAATGCCATTAACCATGCTTGAAGTGAAGAT 720
QY 766 TTATCTTTAGAAAAATGCTATTTGAAAAAAGCTTTTGACAAAGCAATCCGCAATTAATCTGAT 825
| | | | |
Db 721 TTATCTTTAGAAAAATGCTATTTGAAAAAAGCTTTTGACAAAGCAATCCGCAATTAATCTGAT 780
QY 826 CATTGTGGCCACCACTTGAGCCAAAGACTAAATCCTTAGTTCAGATTTTGAAGATTTA 885
| | | | |
Db 781 CATTGTGGCCACCACTTGAGCCAAAGACTAAATCCTTAGTTCAGATTTTGAAGATTTA 840
QY 886 CGAATCTTGCTGAGTATCTCTCACTATGATTTGTGTCATTTCTTAATCTTCTGGA 945
| | | | |
Db 841 CGAATCTTGCTGAGTATCTCTCACTATGATTTGTGTCATTTCTTAATCTTCTGGA 900
QY 946 TCTCTGAGACCAACGAAAAAGCTTTTGTGCAGAAATTCAGTTGGCTGTTCCTGATCC 1005
| | | | |
Db 901 TCTCTGAGACCAACGAAAAAGCTTTTGTGCAGAAATTCAGTTGGCTGTTCCTGATCC 960
QY 1006 AGCACTCTCATGTTTATTAATGCTCGAGCAAGGTTTATCTCTCCAGATGCCAAATG 1065
| | | | |
Db 961 AGCACTCTCATGTTTATTAATGCTCGAGCAAGGTTTATCTCTCCAGATGCCAAATG 1020
QY 1066 AGTAAAAAAGAAAAATATCTGAATAATGAAATTAAGAAGGGAAGAACAACAAAG 1125
| | | | |
Db 1021 AGTAAAAAAGAAAAATATCTGAATAATGAAATTAAGAAGGGAAGAACAACAAAG 1080
QY 1126 GAACGTGCTCTAGAAAGCAACCAAGTGGAGGCACTGACTGAAGTATTAAGAAAT 1185
| | | | |
Db 1081 GAACGTGCTCTAGAAAGCAACCAAGTGGAGGCACTGACTGAAGTATTAAGAAAT 1140
QY 1186 GAGGCAAGAAATAGAGGAGTGAAGCTCTTGAGTGCAGAGTCAAGTACTGATTTGCA 1245
| | | | |
Db 1141 GAGGCAAGAAATAGAGGAGTGAAGCTCTTGAGTGCAGAGTCAAGTACTGATTTGCA 1200
QY 1246 AGTATGATCCGAACATGTTCCAGCTGAGAGTATATCACTCTTGAGGCGAGGCTTC 1305
| | | | |
Db 1201 AGTATGATCCGAACATGTTCCAGCTGAGAGTATATCACTCTTGAGGCGAGGCTTC 1260
QY 1306 TTATTTAGAGCTCTACAGAAACCTTTGAGAGGATAGCAAAAGCTGAAGAGTCTGATG 1365
| | | | |
Db 1261 TTATTTAGAGCTCTACAGAAACCTTTGAGAGGATAGCAAAAGCTGAAGAGTCTGATG 1320
QY 1366 AAATTTAGAGAGCAACCTTCAAGAGAAATTAAGAAATCTCACAACAAACCTTAAGAC 1425
| | | | |
Db 1321 AAATTTAGAGAGCAACCTTCAAGAGAAATTAAGAAATCTCACAACAAACCTTAAGAC 1380
QY 1426 CCCCAGAAACAAAGAGGCGCTTCTACCAAGAAAGAACCTCAAAAAGAAAGAAAGCGAAG 1485
| | | | |
Db 1381 CCCCAGAAACAAAGAGGCGCTTCTACCAAGAAAGAACCTCAAAAAGAAAGAAAGCGAAG 1440

QY 1486 TTGACCTTAACCTCAATGTTAGGAAAACTGAAAGCACTGGAACGAAAGCATGTGAG 1545
| | | | |
Db 1441 TTGACCTTAACCTCAATGTTAGGAAAACTGGAACGAAAGCATGTGAG 1500
QY 1546 GAAGGATATGCTGAGAAATTAAGCAGTAGACCCAGAAAGCTCCCGCAAGAAATTAAGCAT 1605
| | | | |
Db 1501 GAAGGATATGCTGAGAAATTAAGCAGTAGACCCAGAAAGCTCCCGCAAGAAATTAAGCAT 1560
QY 1606 GAAGATTTGATGTAATTTTGTCTATGCGAGTCTCTTTGCGATTCCTCAAGAAACCCCTC 1665
| | | | |
Db 1561 GAAGATTTGATGTAATTTTGTCTATGCGAGTCTCTTTGCGATTCCTCAAGAAACCCCTC 1620
QY 1666 ACATATCATCCATCCGCTTCTGAGGTTGAGGAGACCCCTATCTCTGCAAGAGTACTACAT 1725
| | | | |
Db 1621 ACATATCATCCATCCGCTTCTGAGGTTGAGGAGACCCCTATCTCTGCAAGAGTACTACAT 1680
QY 1726 GAAGTGGAGCCAGAGATACGTGTTCTTTATAGCAGAGCTAACCTTTGTTGCGAGCTT 1785
| | | | |
Db 1681 GAAGTGGAGCCAGAGATACGTGTTCTTTATAGCAGAGCTAACCTTTGTTGCGAGCTT 1740
QY 1786 GAAATTTACAGGCGAGTACGCGCGGGAACCTGTAGGAGTTTACTTTATATACGA 1845
| | | | |
Db 1741 GAAATTTACAGGCGAGTACGCGCGGGAACCTGTAGGAGTTTACTTTATATACGA 1800
QY 1846 GGTTCACATGAGGAACAACCTATCTCACTGCTTTGCGGAACAAAGCAAGCTTTTGAA 1905
| | | | |
Db 1801 GGTTCACATGAGGAACAACCTATCTCACTGCTTTGCGGAACAAAGCAAGCTTTTGAA 1860
QY 1906 AAATCTATTAAGGAAAAAGCAAGCATGTTGTCCCTGAAAGAAAGCAAGCAGATGAA 1965
| | | | |
Db 1861 AAATCTATTAAGGAAAAAGCAAGCATGTTGTCCCTGAAAGAAAGCAAGCAGATGAA 1920
QY 1966 ACAAACTTAGACCTAGTAAGAGGACAGCAATCTGCAATGTTCCATCTGACACTGGAAA 2025
| | | | |
Db 1921 ACAAACTTAGACCTAGTAAGAGGACAGCAATCTGCAATGTTCCATCTGACACTGGAAA 1980
QY 2026 GCGGCTGCCAGGAACAAGATGTTACACAGCAAAAGCTATGTCATATGCGGAATTT 2085
| | | | |
Db 1981 GCGGCTGCCAGGAACAAGATGTTACACAGCAAAAGCTATGTCATATGCGGAATTT 2040
QY 2086 CGAAGTATGCTTCATCTGATTCATCCATGTCGCGGAGCTTGACATTTCAACCCGAGACTTA 2145
| | | | |
Db 2041 CGAAGTATGCTTCATCTGATTCATCCATGTCGCGGAGCTTGACATTTCAACCCGAGACTTA 2100
QY 2146 GAGGTTGAGATTAACATCTCTACCTCCAGAAATGTCGTGAGGCGCAAGAGTATCAGTAT 2205
| | | | |
Db 2101 GAGGTTGAGATTAACATCTCTACCTCCAGAAATGTCGTGAGGCGCAAGAGTATCAGTAT 2160
QY 2206 TTAATCGGCTCTTTAAATTAACGGCGCGCTCTACAGCCAGTGCATCTCATGTCCGCTAC 2265
| | | | |
Db 2161 TTAATCGGCTCTTTAAATTAACGGCGCGCTCTACAGCCAGTGCATCTCATGTCCGCTAC 2220
QY 2266 TACAAGGCTCCCGTCTTGATTAAGTTGAGCTTGAACCTGCAAGGCTTCTCTCTACTTCC 2325
| | | | |
Db 2221 TACAAGGCTCCCGTCTTGATTAAGTTGAGCTTGAACCTGCAAGGCTTCTCTCTACTTCC 2280
QY 2326 CGAGTGCCCTGTTTTCAGAGATCTCCAGCAATGACATTAATGTTCCAAACTACTCTTCTT 2385
| | | | |
Db 2281 CGAGTGCCCTGTTTTCAGAGATCTCCAGCAATGACATTAATGTTCCAAACTACTCTTCTT 2340
QY 2386 ACACTTACTTCCCGACACTACGAGATTCCTGTGAGGCGGCTCTCTCTATGCAAGGCGGAG 2445
| | | | |
Db 2341 ACACTTACTTCCCGACACTACGAGATTCCTGTGAGGCGGCTCTCTCTATGCAAGGCGGAG 2400
QY 2446 TTGTTTGAAGAGCTGAACAAAGCAAGCAACAGCTGATCGGCGAGCAGCACTGCGCAT 2505
| | | | |
Db 2401 TTGTTTGAAGAGCTGAACAAAGCAAGCAACAGCTGATCGGCGAGCAGCACTGCGCAT 2460
QY 2506 ACAGCAGATTTCTGAAACCTTCCGAGTCTAGAGAAATTAATCTCTCTCCCAAGACTTC 2565
| | | | |
Db 2461 ACAGCAGATTTCTGAAACCTTCCGAGTCTAGAGAAATTAATCTCTCTCCCAAGACTTC 2520
QY 2566 TTGTTTAAAAATGCCAGGCGGATGCCAAAAACTGCCGCTCTTTGATGCAACAGCTTAAG 2625

Db	2521	TTGTTAAATAATGCCAGGGGGAATGGCAAAAACTGCGGTCTCTTATATCACCACCTTTAG	2580
Qy	2626	AACATGCGACAATTTACGAGCCCTGTGCACAAGAGCTCAGAGTATTTCTGGGAAATGCT	2685
Db	2581	AACATGCGACAATTTACGAGCCCTGTGCACAAGAGAGCTCAGAGATTTCTGGGAAATGCT	2640
Qy	2686	GCAAAATGCCAAACAGCTTATGATTTCAATTCACACCTCTTTGGCAAGAGTGCTATCAAAA	2745
Db	2641	GCAAAATGCCAAACAGCTTATGATTTCAATTCACACCTCTTTGGCAAGAGTGCTATCAAAA	2700
Qy	2746	GGAAAAAGGAAAAAGTGA	2763
Db	2701	GGAAAAAGGAAAAAGTGA	2718

RESULT 2

```

: Sequence 14: Application US/09650855
: Patient No. 6365355
:
: GENERAL INFORMATION:
:
: APPLICANT: MCGUTHER-MALONEY, SANDRA
: APPLICANT: LAWRENCE LIVERMORE NATIONAL LABORATORY
: TITLE OF INVENTION: CHIMERIC PROTEINS FOR DETECTION AND QUANTITATION OF DNA
: TITLE OF INVENTION: MUTATIONS, DNA SEQUENCE VARIATIONS, DNA DAMAGE AND DNA
: TITLE OF INVENTION: MISMATCHES
: FILE REFERENCE: IL-10284
: CURRENT APPLICATION NUMBER: US/09/650,855
: CURRENT FILING DATE: 2000-08-29
: PRIOR APPLICATION NUMBER: 60/192,764
: PRIOR FILING DATE: 2000-03-28
: NUMBER OF SEQ ID NOS: 106
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 14
:
: LENGTH: 2718
:
: TYPE: DNA
:
: ORGANISM: Homo sapiens
: S-09-650-855-14

```

Query Match	Score	DB 4	Length
70.5%	2718		2718

```
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2718; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

QY	46	ATGGCGCGCGCTGCGAGAGTACAGACGACACTGGTGGTCTGGAACTGCTCGACACTGACGGG	105
Db	1	ATGGCGCGCGCTGCTGGAGTACAGACGACACTGGTGGTGGAACTGCTCGACACTGACGGG	60
QY	106	CTATAGTGTGCGCGCCCGGGGCTGGGGCGGACCGGCGCTGCTGCAACTTCTCCACCTG	165
Db	61	CTATAGTGTGCGCGCCCGGGGCTGGGGCGGACCGGCGCTGCTGCAACTTCTCCACCTG	120
QY	166	CACGTGCACCCAGCCCTGCTGTGCTGTGCTCAACAAGCAGCGCGGCGGAGAGAGTAT	225
Db	121	CACGTGCACCCAGCCCTGCTGTGCTGTGCTCAACAAGCAGCGCGGCGGAGAGAGTAT	180
QY	226	TTTATTCATCAGCTGAAGATGAAAGAGTTGAACACCTCCCTCGCGCTGTAACAAATGAA	285
Db	181	TTTATTCATCAGCTGAAGATGAAAGAGTTGAACACCTCCCTCGCGCTGTAACAAATGAA	240
QY	286	ATCCAGCAACACCTGCCCTATGAGAGTTTACACACCAAGTGGTGTATATTTGGACAAGT	345
Db	241	ATCCAGCAACACCACTGCCCTATGAGAGTTTACACACCAAGTGGTGTATATTTGGACAAGT	300
QY	346	AGCATCTTGTGGTGTGACTCTTACTGATAGAAATACCTTCAGATTAAATTAATGCGCATC	405
Db	301	AGCATCTTGTGGTGTGACTCTTACTGATAGAAATACCTTCAGATTAAATTAATGCGCATC	360
QY	406	TTGCTGTATAGAGCCCAAGAAATATATGAGTCTTGTCAAGAAGCATTCATCTTGCGCTC	465
Db	361	TTGCTGTATAGAGCCCAAGAAATATATGAGTCTTGTCAAGAAGCATTCATCTTGCGCTC	420
QY	466	TTTGGCCAGAAAAACAAGCTGTTTTATTAAAGCTTTCACAGACAATGCTGTTGCCCTT	525

D	b	421	TTTGCCAGAAAAAACAACGCTGGTTTATTAAAGCTTTCACAGACATCCGTTGCTTT	480
O	y	526	GATACGTGTTTTGTCACTGTGGAAAGAGTATGAGAAATCTTTTGTGAGAACTGTAT	585
D	b	481	GATACGTGTTTTGTCACTGTGGAAAGAGTATGAGAAATCTTTTGTGAGAACTGTAT	540
O	y	586	CTGTGCGCAAGGTTCCATGTAGCAGTAACTCATTTTGTAGAACAGCAACAACCTGAAGT	645
D	b	541	CTGTGCGCAAGGTTCCATGTAGCAGTAACTCATTTTGTAGAACAGCAACAACCTGAAGT	600
O	y	646	GTAGAAATCCATGTTTCTATGACACCACATGCTGTATACAGCTCCTATCTAGGAC	705
D	b	601	GTAGAAATCCATGTTTCTATGACACCACATGCTGTATACAGCTCCTATCTAGGAC	660
O	y	706	ATTTTAAATGCATGTCTAAAGAACTAAATGCGCATACCCATCGCTTAAGTGGAGAT	765
D	b	661	ATTTTAAATGCATGTCTAAAGAACTAAATGCGCATACCCATCGCTTAAGTGGAGAT	720
O	y	766	TTATCTTTAGAAAAATGCTATTGGAAAACTTTTGTAGAACAAATCCGCCATTATCTGAT	825
D	b	721	TTATCTTTAGAAAAATGCTATTGGAAAACTTTTGTAGAACAAATCCGCCATTATCTGAT	780
O	y	826	CCTTTGTGGCACCGACTTGGAGCCAGACATAATCCTTAGTTCAGATTTGAAGATATTA	885
D	b	781	CCTTTGTGGCACCGACTTGGAGCCAGACATAATCCTTAGTTCAGATTTGAAGATATTA	840
O	y	886	CGAATTTTGCGCAGTATGCTCTCAGTATGATTTGTGCATATTTCTTAATCTCTGGAA	945
D	b	841	CGAATTTTGCGCAGTATGCTCTCAGTATGATTTGTGCATATTTCTTAATCTCTGGAA	900
O	y	946	TCTCTGAGAGCAACGAAAAAGCTTTTGTGCAGAAATTCAGTTGCGCTTTCTTACTCC	1005
D	b	901	TCTCTGAGAGCAACGAAAAAGCTTTTGTGCAGAAATTCAGTTGCGCTTTCTTACTCC	960
O	y	1006	AGCACCCTCGATGTTTATAAATGCTCGACAGCGTTTATCATCTTCCAGATGCCAAAATG	1065
D	b	961	AGCACCCTCGATGTTTATAAATGCTCGACAGCGTTTATCATCTTCCAGATGCCAAAATG	1020
O	y	1066	AGTAAAAAGAAAAATATCTGAAAAAATGGAATTAAGAAGGGGAACAAACAAAAG	1125
D	b	1021	AGTAAAAAGAAAAATATCTGAAAAAATGGAATTAAGAAGGGGAACAAACAAAAG	1080
O	y	1126	GAACTGTGCTCAGAAAGCAACCCAAAAGTGGAGGCACTCACTGAAGTATTAAAGAAATY	1185
D	b	1081	GAACTGTGCTCAGAAAGCAACCCAAAAGTGGAGGCACTCACTGAAGTATTAAAGAAATY	1140
O	y	1186	GAGCGAGAAAAATAAGAGAGTGAAGCTCTTGTTGTTCCAGGCTCAGTACTGTTTGTGCA	1245
D	b	1141	GAGCGAGAAAAATAAGAGAGTGAAGCTCTTGTTGTTCCAGGCTCAGTACTGTTTGTGCA	1200
O	y	1246	AGTGTATGCCGAACTATGTTCCCAAGCTGAGAGCATATATCACTCTTGTGAGCGGAGCTTC	1305
D	b	1201	AGTGTATGCCGAACTATGTTCCCAAGCTGAGAGCATATATCACTCTTGTGAGCGGAGCTTC	1260
O	y	1306	TTATTTAGAGGCTCTACAGAGAAACCTTTGAAGAGATAGCAAAAGCTGAAGATCTGGATG	1365
D	b	1261	TTATTTAGAGGCTCTACAGAGAAACCTTTGAAGAGATAGCAAAAGCTGAAGATCTGGATG	1320
O	y	1366	AAATTTAGGAAGGAAGACAGTTCAAGAGAAATTAGGAATCTCACAAAAGACCTAAAGAC	1425
D	b	1321	AAATTTAGGAAGGAAGACAGTTCAAGAGAAATTAGGAATCTCACAAAAGACCTAAAGAC	1380
O	y	1426	CCCCAAAACAAAGAACGGGCTTTACCAAAAGAAACCCCTCAAAAAGAAAAACGGGAAG	1485
D	b	1381	CCCCAAAACAAAGAACGGGCTTTACCAAAAGAAACCCCTCAAAAAGAAAAACGGGAAG	1440
O	y	1486	TTGACCTTAACTCAAAATGTTAGGAAAAACCTGGAAGACTGGAAGAGAGAGATCTCGAG	1545
D	b	1441	TTGACCTTAACTCAAAATGTTAGGAAAAACCTGGAAGACTGGAAGAGAGAGATCTCGAG	1500
O	y	1546	GAAAGATATCTGTGAGAAATTAAGCAGTAGCCAGAAAGCTGCCGGAAGAAATTAAGCAT	1605
D	b	1501	GAAAGATATCTGTGAGAAATTAAGCAGTAGCCAGAAAGCTGCCGGAAGAAATTAAGCAT	1560

QY 1606 GAAGAATTTGATGTAATTTGTCATGAGTGCCTTTTGGAGATCTGAAAGACCCCTC 1665
DB 1561 GAAGAATTTGATGTAATTTGTCATGAGTGCCTTTTGGAGATCTGAAAGACCCCTC 1620
QY 1666 ACTATCATCCATCCGCTTCTGGGTGGACGACCCCTATGCTGTGCAAGAGGTACTACAT 1725
DB 1621 ACTATCATCCATCCGCTTCTGGGTGGACGACCCCTATGCTGTGCAAGAGGTACTACAT 1680
QY 1726 GAAGTGGACCCAAAGATACCTGCTTATAGACGACGACCTTAACCTTTTGGGAGCTT 1785
DB 1681 GAAGTGGACCCAAAGATACCTGCTTATAGACGACGACCTTAACCTTTTGGGAGCTT 1740
QY 1786 GAAATTTACAGGGCGAGTAGGCGCTGGGAAACCTCTGAGGCTTACTTTCTTATATACGA 1845
DB 1741 GAAATTTACAGGGCGAGTAGGCGCTGGGAAACCTCTGAGGCTTACTTTCTTATATACGA 1800
QY 1846 GGTTCACACTGAGACACAGCCTATCTCAGCTGCTTTGCGGAAAGAAAGAGAGCTTTTGA 1905
DB 1801 GGTTCACACTGAGACACAGCCTATCTCAGCTGCTTTGCGGAAAGAAAGAGAGCTTTTGA 1860
QY 1906 AATCTATAGGAGAAAGACAGCATGCTTCCCTGAAGAAAGAGAGAGAGATGAA 1965
DB 1861 AATCTATAGGAGAAAGACAGCATGCTTCCCTGAAGAAAGAGAGAGAGATGAA 1920
QY 1966 ACAAACTTAGACTAGTAGAGGACAGCATCTGAGATGTTTCCACTGACACTTCGGA 2025
DB 1921 ACAAACTTAGACTAGTAGAGGACAGCATCTGAGATGTTTCCACTGACACTTCGGA 1980
QY 2026 GCCGTGGCCAGACAGACAGATGTCACAGCAAGACATAGTTGATGCTGAATTT 2085
DB 1981 GCCGTGGCCAGACAGACAGATGTCACAGCAAGACATAGTTGATGCTGAATTT 2040
QY 2086 CGAAGTGGAGCTTCCATCTGATCCATCGTCGCGGCGATTGACATTAACCCGAGCTT 2145
DB 2041 CGAAGTGGAGCTTCCATCTGATCCATCGTCGCGGCGATTGACATTAACCCGAGCTT 2100
QY 2146 GAGGTGGAGATTAACATCTCTCACTCAGAAATGTCGTGGAGCGCAGAGATACATGAT 2205
DB 2101 GAGGTGGAGATTAACATCTCTCACTCAGAAATGTCGTGGAGCGCAGAGATACATGAT 2160
QY 2206 TTTAATGGGCTTTAAATAAAGCGCGCTCTTAAGCAGTGCATCTCCATGCTCCCTAC 2265
DB 2161 TTTAATGGGCTTTAAATAAAGCGCGCTCTTAAGCAGTGCATCTCCATGCTCCCTAC 2220
QY 2266 TTAACAGGTCCTGCTTGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 2325
DB 2221 TTAACAGGTCCTGCTTGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 2280
QY 2326 CGAAGTGGCTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2385
DB 2281 CGAAGTGGCTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2340
QY 2386 ACACCTTCACTTCCCGAGACTACGAGATTTCTGTGCTCCCTCTCTCATGCAAGCGGAG 2445
DB 2341 ACACCTTCACTTCCCGAGACTACGAGATTTCTGTGCTCCCTCTCTCATGCAAGCGGAG 2400
QY 2446 TTTGTTGAGAGCTGAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 2505
DB 2401 TTTGTTGAGAGCTGAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 2460
QY 2506 ACAGCAGATTTCTGAAACCTTTCCGAGTGAAGATATATATCTGCTCCCAAGACTTC 2565
DB 2461 ACAGCAGATTTCTGAAACCTTTCCGAGTGAAGATATATATCTGCTCCCAAGACTTC 2520
QY 2566 TTTGTTAAATATGCGAGGGTGAATGCCAAAACTGCGCTCTTATGATGACCAAGTTAAG 2625
DB 2521 TTTGTTAAATATGCGAGGGTGAATGCCAAAACTGCGCTCTTATGATGACCAAGTTAAG 2580
QY 2626 AATATGCGAATTTAGACGCTCTGACAGAGAGAGATGATGATGATGATGATGATGATGAT 2685
DB 2581 AATATGCGAATTTAGACGCTCTGACAGAGAGAGATGATGATGATGATGATGATGATGAT 2640

QY 2686 GCAAATGCCAAACAGCTTTATGATTTATTCATTCACACCTCTTTTGCAGAGTGTATCAAAA 2745
DB 2641 GCAAATGCCAAACAGCTTTATGATTTATTCATTCACACCTCTTTTGCAGAGTGTATCAAAA 2700
QY 2746 GCAAAAGGCAAAAAGTGA 2763
DB 2701 GCAAAAGGCAAAAAGTGA 2718

RESULT 3
US-09-651-656-12
; Sequence 12, Application US/09651656
; Patent No. 6340566
; GENERAL INFORMATION:
; APPLICANT: MCUTHER-MALONEY, SANDRA
; TITLE OF INVENTION: LAWRENCE LIVERMORE NATIONAL LABORATORY
; TITLE OF INVENTION: DETECTION AND QUANTIFICATION OF SINGLE NUCLEOTIDE
; TITLE OF INVENTION: POLYMORPHISMS, DNA SEQUENCE VARIATIONS, DNA MUTATIONS,
; FILE REFERENCE: IL-10689
; CURRENT APPLICATION NUMBER: US/09/651,656
; CURRENT FILING DATE: 2000-08-29
; PRIOR APPLICATION NUMBER: 60/192,764
; NUMBER OF SEQ ID NOS: 106
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 12
; LENGTH: 1101
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-651-656-12

Query Match 28.5%; Score 1101; DB 4; Length 1101;
Best Local Similarity 100.0%; Pred. No. 1.9e-284;
Matches 1101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 46 ATGGCGCGCTGCTGGAGTACAGAGCAGACAGCTGCTGTAACCTCTCCACACTGACGGG 105
DB 1 ATGGCGCGCTGCTGGAGTACAGAGCAGACAGCTGCTGTAACCTCTCCACACTGACGGG 60
QY 106 CTAGTAGTGTGCGCGCGCGGCGCTGCGCGGAGCGAGCGGCTCTTACCACTTCTCCAGCTG 165
DB 61 CTAGTAGTGTGCGCGCGCGGCGCTGCGCGGAGCGAGCGGCTCTTACCACTTCTCCAGCTG 120
QY 166 CACTGCAACCCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 225
DB 121 CACTGCAACCCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 180
QY 226 TTTATCATATAGCTGAAGATAGAGAGTGAACACCTCCCTGCGCGTGAACAAATGAA 285
DB 181 TTTATCATATAGCTGAAGATAGAGAGTGAACACCTCCCTGCGCGTGAACAAATGAA 240
QY 286 ATCAACAAGCAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 345
DB 241 ATCAACAAGCAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 300
QY 346 AGGATACCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 405
DB 301 AGGATACCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 360
QY 406 TTTGCTATATAGAGCCCAAGATATATGAGTCTTGTCAAGAGACATTCATCTGGCCTC 465
DB 361 TTTGCTATATAGAGCCCAAGATATATGAGTCTTGTCAAGAGACATTCATCTGGCCTC 420
QY 466 TTTGCGCAAGAAACCAAGCTGTTTATTAAGCTTTTACACACAAATGCTGTGCTTT 525
DB 421 TTTGCGCAAGAAACCAAGCTGTTTATTAAGCTTTTACACACAAATGCTGTGCTTT 480
QY 526 GATACGCTTTTGTGCTATGGAAGAGATGAGAAATCTTTTGTGAGAGAACTGTAT 585
DB 481 GATACGCTTTTGTGCTATGGAAGAGATGAGAAATCTTTTGTGAGAGAACTGTAT 540
QY 586 CTGTGGCAAGGTTCCATGTAGCAGTAACTCAATTTTGAACAGCAACAACTGAAGTT 645


```
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; JTELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: PTZgpt-F15
; US-08-232-463-14

Query Match          2.0%; Score 78; DB 1; Length 7218;
Best Local Similarity 3.8%; Pred. No. 1,4e-10;
Matches 15; Conservative 242; Mismatches 137; Indels 0; Gaps 0;

QY 1035 AAGGTTTATCATCTTCGAGATGCCAAATGATAAAGAAAGAAATATCTGAAAT 1094
    ||| |||
DB 1447 AACAATTTGCTACRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1388
QY 1095 GGAATTAAGAAGGAGCAACAAGAAAGAACTGCTCTGAGAAAGCAACCAAGT 1154
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1387 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1328
QY 1155 GGAGGCACTGACTGAAGTATTAAGAAATTGAGCAAGAAATAGAGAGTGAAGCT 1214
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1327 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1268
QY 1215 TGCTGTCGACGTCAGTACTGATTTGTGCAAGTGAACCAACTGTCACAGCTGAG 1274
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1267 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1208
QY 1275 AGACTATATCACTCTTGAGCGAGGCTCTTATTGAGGCTTACAGAAACCTTTGA 1334
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1207 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1148
QY 1335 GAAGATAGCAAGTGAAGAACTGATGAATTTAGAGAGGAAGAGCTTCAAGAG 1394
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

```
DB 1147 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1088
QY 1395 AATTAGAAATCTCAACAAAGACCTAAGAGCCCC 1428
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1087 RRRRRRRRRRRRRRRRRRRRRRRRRRATCGCAAGCTCCC 1054

RESULT 6
US-09-078-294-4/C
; Sequence 4, Application US/09078294
; Patent No. 6255211
; GENERAL INFORMATION:
; APPLICANT: Choo, Kong-Hong Andy
; APPLICANT: Du Sart, Desiree
; APPLICANT: Cancilla, Michael R.
; TITLE OF INVENTION: A NOVEL NUCLEIC ACID MOLECULE
; FILE REFERENCE: Davies Col
; CURRENT APPLICATION NUMBER: US/09/078,294
; CURRENT FILING DATE: 1998-05-13
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 4
; LENGTH: 80246
; TYPE: DNA
; ORGANISM: Nucleotide sequence of NC-contig
US-09-078-294-4

Query Match          1.8%; Score 68.6; DB 4; Length 80246;
Best Local Similarity 70.2%; Pred. No. 1.7e-07;
Matches 92; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

QY 3723 CAGCATCATATCTCTAGAAATACCTAATGGAGTGAGCCGAGATCGACACTGC 3782
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 4751 CAGGAGATCATCTTGAACCCAGGAGCGAAGTTGAGTGACCGAGATCGCCACTAC 4692
QY 3783 ACCCTGCTGGCGGACAGAGTGAAGTGTCTTATTACAAAAGAAAGAAAGAA 3842
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 4691 ACTCCAGCCCTGGGTGACAGAGCACTGTCTAAAAAAGAAAAAAGAAAAA 4632
QY 3843 TACACCTAAG 3853
DB 4631 AAAAAAAG 4621

RESULT 7
US-09-345-882-1
; Sequence 1, Application US/09345882
; Patent No. 6399373
; GENERAL INFORMATION:
; APPLICANT: Bougueteloret, Lydie
; TITLE OF INVENTION: A NUCLEIC ACID ENCODING A RETINOBLASTOMA BINDING PROTEIN (RBP-
; FILE REFERENCE: GENSET 031A
; CURRENT APPLICATION NUMBER: US/09/345,882
; CURRENT FILING DATE: 1999-06-30
; PRIOR APPLICATION NUMBER: US 60/091,315
; PRIOR FILING DATE: 1998-06-30
; PRIOR APPLICATION NUMBER: US 60/111,909
; PRIOR FILING DATE: 1998-12-10
; NUMBER OF SEQ ID NOS: 140
; SOFTWARE: Patent.pm
; SEQ ID NO 1
; LENGTH: 162450
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: allele
; LOCATION: 72794
; OTHER INFORMATION: 5-124-273 : polymorphic base A or G
; FEATURE:
; NAME/KEY: allele
; LOCATION: 88073
; OTHER INFORMATION: 5-127-261 : polymorphic base A or C
```


FEATURE:
NAME/KEY: allele
LOCATION: 90842
OTHER INFORMATION: 99-1437-325 : polymorphic base A or G
FEATURE:
NAME/KEY: allele
LOCATION: 93714
OTHER INFORMATION: 5-128-60 : polymorphic base deletion of GT
FEATURE:
NAME/KEY: allele
LOCATION: 97122
OTHER INFORMATION: 99-1442-224 : polymorphic base G or T
FEATURE:
NAME/KEY: allele
LOCATION: 97152
OTHER INFORMATION: 5-129-144 : polymorphic base deletion of T
FEATURE:
NAME/KEY: allele
LOCATION: 99098
OTHER INFORMATION: 5-130-257 : polymorphic base A or G
FEATURE:
NAME/KEY: allele
LOCATION: 99117
OTHER INFORMATION: 5-130-276 : polymorphic base A or G
FEATURE:
NAME/KEY: allele
LOCATION: 103806
OTHER INFORMATION: 5-131-395 : polymorphic base A or T
FEATURE:
NAME/KEY: allele
LOCATION: 106940
OTHER INFORMATION: 5-133-375 : polymorphic base insertion of A
FEATURE:
NAME/KEY: allele
LOCATION: 108106
OTHER INFORMATION: 5-135-155 : polymorphic base insertion of A
FEATURE:
NAME/KEY: allele
LOCATION: 108149
OTHER INFORMATION: 5-135-198 : polymorphic base insertion of GTTT
FEATURE:
NAME/KEY: allele
LOCATION: 108308
OTHER INFORMATION: 5-135-357 : polymorphic base A or G
FEATURE:
NAME/KEY: allele
LOCATION: 108471
OTHER INFORMATION: 5-136-174 : polymorphic base C or T
FEATURE:
NAME/KEY: allele
LOCATION: 134134
OTHER INFORMATION: 5-140-120 : polymorphic base C or T
FEATURE:
NAME/KEY: allele
LOCATION: 134362
OTHER INFORMATION: 5-140-348 : polymorphic base insertion of A
FEATURE:
NAME/KEY: allele
LOCATION: 134374
OTHER INFORMATION: 5-140-361 : polymorphic base insertion of CA
FEATURE:
NAME/KEY: allele
LOCATION: 146328
OTHER INFORMATION: 5-143-84 : polymorphic base A or G
FEATURE:
NAME/KEY: allele
LOCATION: 146345
OTHER INFORMATION: 5-143-101 : polymorphic base A or C
FEATURE:
NAME/KEY: allele
LOCATION: 150329
OTHER INFORMATION: 5-145-24 : polymorphic base A or G
FEATURE:

NAME/KEY: allele
LOCATION: 160031
OTHER INFORMATION: 5-148-352 : polymorphic base G or T
FEATURE:
NAME/KEY: allele
LOCATION: 72771..72817
OTHER INFORMATION: polymorphic fragment 5-124-273 SEQ ID30
FEATURE:
NAME/KEY: allele
LOCATION: 72771..72817
OTHER INFORMATION: polymorphic fragment 5-124-273 SEQ ID51
FEATURE:
NAME/KEY: allele
LOCATION: 88050..88096
OTHER INFORMATION: polymorphic fragment 5-127-261 SEQ ID31
FEATURE:
NAME/KEY: allele
LOCATION: 88050..88096
OTHER INFORMATION: polymorphic fragment 5-127-261 SEQ ID52
FEATURE:
NAME/KEY: allele
LOCATION: 90819..90865
OTHER INFORMATION: complement polymorphic fragment 99-1437-325 SEQ ID49
FEATURE:
NAME/KEY: allele
LOCATION: 90819..90865
OTHER INFORMATION: complement polymorphic fragment 99-1437-325 SEQ ID70
FEATURE:
NAME/KEY: allele
LOCATION: 93690..93736
OTHER INFORMATION: polymorphic fragment 5-128-60 SEQ ID53
FEATURE:
NAME/KEY: allele
LOCATION: 93690..93736
OTHER INFORMATION: polymorphic fragment 5-128-60 SEQ ID50
FEATURE:
NAME/KEY: allele
LOCATION: 97099..97145
OTHER INFORMATION: polymorphic fragment 99-1442-224 SEQ ID71
FEATURE:
NAME/KEY: allele
LOCATION: 97099..97145
OTHER INFORMATION: polymorphic fragment 5-129-144 SEQ ID33
FEATURE:
NAME/KEY: allele
LOCATION: 97130..97177
OTHER INFORMATION: polymorphic fragment 5-129-144 SEQ ID54
FEATURE:
NAME/KEY: allele
LOCATION: 97130..97177
OTHER INFORMATION: polymorphic fragment 5-129-144 SEQ ID55
FEATURE:
NAME/KEY: allele
LOCATION: 99075..99121
OTHER INFORMATION: polymorphic fragment 5-130-257 SEQ ID34
FEATURE:
NAME/KEY: allele
LOCATION: 99075..99121
OTHER INFORMATION: polymorphic fragment 5-130-257 SEQ ID55
FEATURE:
NAME/KEY: allele
LOCATION: 99094..99140
OTHER INFORMATION: polymorphic fragment 5-130-276 SEQ ID35
FEATURE:
NAME/KEY: allele
LOCATION: 99094..99140
OTHER INFORMATION: polymorphic fragment 5-130-276 SEQ ID56
FEATURE:
NAME/KEY: allele
LOCATION: 103783..103828
OTHER INFORMATION: polymorphic fragment 5-131-395 SEQ ID36
FEATURE:
NAME/KEY: allele

```

LOCATION: 103783..103828
OTHER INFORMATION: polymorphic fragment 5-131-395 SEQ ID57
FEATURE:
NAME/KEY: allele
LOCATION: 106918..106966
OTHER INFORMATION: polymorphic fragment 5-133-375 SEQ ID57
FEATURE:
NAME/KEY: allele
LOCATION: 106918..106966
OTHER INFORMATION: polymorphic fragment 5-133-375 SEQ ID58
FEATURE:
NAME/KEY: allele
LOCATION: 108084..108130
OTHER INFORMATION: polymorphic fragment 5-135-155 SEQ ID58
FEATURE:
NAME/KEY: allele
LOCATION: 108084..108130
OTHER INFORMATION: polymorphic fragment 5-135-155 SEQ ID59
FEATURE:
NAME/KEY: allele
LOCATION: 108127..108177
OTHER INFORMATION: polymorphic fragment 5-135-198 SEQ ID59
FEATURE:
NAME/KEY: allele
LOCATION: 108127..108177
OTHER INFORMATION: polymorphic fragment 5-135-198 SEQ ID60
FEATURE:

```

Query Match	1.8%	Score	68.2	DB	4	Length	162450
Best Local	Similarly	Pred	No. 3.2e-07				
Matches	82	Conservative					
		Mismatches	23			Indels	0
						Gaps	0

Oy 3748 ACCTAATTGGCAGTGTGACCCGAGNTGCACCACTGCACCCCCTGCCCTTG66CGACAGATGTGAG 3807
||| ||||||||| ||||||||| ||||||||| ||||||||| ||
Db 99615 ACGGAGCTTTGCAGTAGCACCAAGATGTGGCCGA CTGCAC TCCAGCC TTGGGCGACAGAGCCAG 9967

OY	3808	ACTTGTCTCTATTACAAAAAGAAAGAATACCACTTA	3852
Db	99675	ACTCTGTCCTCAAAAAAAAAAAGCAGAACAAAAACA	99719

```

RESULT 8
US-08-965-048--5
Sequence 5, Application US/08965048
Patent No. 6323244
GENERAL INFORMATION:
APPLICANT: Chen, Hong
APPLICANT: Freilmer, Nelson
TITLE OF INVENTION: METHODS AND COMPOSITIONS AND
TITLE OF INVENTION: TREATMENT OF NEUROPSYCHIATRIC DISORDERS
FILE REFERENCE: 7853-03
CURRENT APPLICATION NUMBER: US/08/965,048
CURRENT FILING DATE: 1997-11-05
NUMBER OF SEQ ID NOS: 8
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 5
LENGTH: 45716
TYPE: DNA
ORGANISM: Homo sapiens
US-08-965-048--5

```

Query Match	1.8%	Score 68	DB 4	Length 45716
Best Local Similarly	83.7%	Pred. No. 1.8e-07		
Matches 77	Conservative	0	Mismatches 15	Indels 0
				Gaps 0

Oy	3755	TGGCAGTGTGACCCGATGTGCACCCTGCCCTGGCGCAGAGTGAGACTTTTGT	3814
Db	21289	TTGCAGTGTGACCCGATGTGCACCCTGCCCTGGCGCAGAGTGAGACTCTGT	21348

Qy	3815	CTCTATTACAAAAAGAAAGAAATACA	3846
Db	21349	CTCAAAACAAAAAATAAAAAAAGCA	21380

```

RESULT 9
US-08-965-048-6
: Sequence 6, Application US/08965048
: Patent No. 6323244
: GENERAL INFORMATION:
: APPLICANT: Chen, Hong
: APPLICANT: Fretmer, Nelson
: TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE DIAGNOSIS AND
: TITLE OF INVENTION: TREATMENT OF NEUROPSYCHIATRIC DISORDERS
: FILE REFERENCE: 7853-093
: CURRENT APPLICATION NUMBER: US/08/965,048
: CURRENT FILING DATE: 1997-11-05
: NUMBER OF SEQ ID NOS: 8
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 6
: LENGTH: 45989
: TYPE: DNA
: ORGANISM: Homo sapiens
US-08-965-048-6

```

Query Match	1.8%	Score 68	DB 4	Length 45989
Best Local Similarity	83.7%	Pred. No. 1.8e-07		
Matches 77	Conservative	0	Mismatches 15	Indels 0
			Gaps	0

Oy 3755 TGGCAGTGAACCGAGATCCGACCACCTGCACCCCCTGGCTGGCGGACAGAGTGAGACTTTGT 3814
| | | | | | | | | | | | | | | | | | | | | |
Db 21403 TTGCAGTGAACCGAGATCCGACCACCTGCACCTTCACGCCCTGGGCAACAGAGTGAGACTCTGT 21466

Qy	3815	CTCTATTACAAAAAGAAAAGAAATACA	3846
Db	21463	CTCAAAACAAAAAAGAAAAAAGAAAAAGA	21494

```

RESULT 10
US-09-426-290-1/c
; Sequence 1, Application US/09426290
; Patent No. 6410712
; GENERAL INFORMATION:
; APPLICANT: Berglynd Ran Olafsdottir
; TITLE OF INVENTION: HUAN NARCOLEPSY GENE
; FILE REFERENCE: 2345.2001-000
; CURRENT APPLICATION NUMBER: US/09/426,290
; CURRENT FILING DATE: 1999-10-25
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: FASTSEQ For Windows Version 4.0
; SEQ ID NO 1

```

```

1 ORGANISM: Homo Sapiens
2
3 FEATURE:
4
5 NAME/KEY: CDS
6
7 LOCATION: (21181)...(21403)
8
9 NAME/KEY: CDS
10
11 LOCATION: (95252)...(95430)
12
13 NAME/KEY: CDS
14
15 LOCATION: (101753)...(101996)
16
17 NAME/KEY: CDS
18
19 LOCATION: (110324)...(110439)
20
21 NAME/KEY: CDS
22
23 LOCATION: (124058)...(124278)
24
25 NAME/KEY: CDS
26
27 LOCATION: (127009)...(127130)
28
29 NAME/KEY: CDS
30
31 LOCATION: (128910)...(129139)
32
33 US-09-426-290-1

```

Query Match	1.88;	Score 67.8;	DB 4;	Length 168575;
Best Local Similarity	78.68;	Pred. No. 4.2e-07;		
Matches 81; Conservative	0;	Mismatches 22;	Indels 0;	Gaps 0;

3755 TGGCAGTGAGCCGAGATCGCACCACTGCACCCCTGCCCTGGGGCAGACGAGTGAGACTTGT 3814

Db 86451 TTGCAGTAGCGGAGATGCTGCCACTGTGCTCCAGCCTGGGCGAGAGAGTGAAGTCCGT 86392

Qy 3815 CTCATTATACAAAAGAAAAGAAATACAACTTAAGTCA 3857

Db 86391 CTCAAAAAAGAAAAGAAAAGAAAATTTACCGCTATGACACA 86349

RESULT 11

US-09-078-294-3/c
Sequence 3, Application US/09078294
Patent No. 6265211

GENERAL INFORMATION:

APPLICANT: Choo, Kong-Hong Andy

APPLICANT: Du Salt, Desiree

APPLICANT: Cancilla, Michael R.

TITLE OF INVENTION: A NOVEL NUCLEIC ACID MOLECULE

FILE REFERENCE: Davies Col

CURRENT APPLICATION NUMBER: US/09/078,294

CURRENT FILING DATE: 1998-05-13

NUMBER OF SEQ ID NOS: 29

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 3

LENGTH: 80595

TYPE: DNA

ORGANISM: Nucleotide sequence of HC-contig

US-09-078-294-3

Query Match 1.7%; Score 67.2; DB 4; Length 80595;
Best Local Similarity 72.5%; Pred. No. 4.1e-07;
Matches 87; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

Qy 3723 CAGCATATATCTCTAGAAATACAACTAATGCGAGTGGCGAGATGCCAGCACCTGC 3782
Db 5032 CAGGGAATCACTTGAAACCCAGGAGCGAAGTTGCAAGTGGAGATCGCGCACTGC 4973

Qy 3783 ACCCTGCTGGGCGAGAGTGAAGCTTGTCTCTATTACAAAAGAAAAGAAA 3842
Db 4972 ACTCAGGCTGGTGCAGAGAGGAGACACTGTCTCAAAAAA 4913

RESULT 12

US-09-128-155-17/c

Sequence 17, Application US/09128155

Patent No. 6117654

GENERAL INFORMATION:

APPLICANT: Pan, Yang

TITLE OF INVENTION: NOVEL MOLECULES OF TANGO-77 RELATED PROTEIN FAMILY

FILE REFERENCE: 09404/052001

CURRENT APPLICATION NUMBER: US/09/128,155

CURRENT FILING DATE: 1998-08-03

EARLIER APPLICATION NUMBER: US 60/091,650

EARLIER FILING DATE: 1998-07-02

EARLIER APPLICATION NUMBER: US 60/054,646

EARLIER FILING DATE: 1997-08-04

NUMBER OF SEQ ID NOS: 18

SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 17

LENGTH: 176373

TYPE: DNA

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: misc_feature

LOCATION: (1)...(176373)

OTHER INFORMATION: n = A,T,C or G

Query Match 1.7%; Score 67; DB 3; Length 176373;
Best Local Similarity 79.8%; Pred. No. 7e-07;
Matches 79; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

Qy 3752 AATTGGCAATGAGCGAGATGCACCACTGCACCCCTGCTGGGCGAGAGTGAAGTCTT 3811

Db 30773 AACTTGGCAATGAGCGAGATGCACCACTGCACCCCTGCTGGGCGAGAGTGAAGTCTT 30714

Qy 3812 TGCTCTATTACAAAAGAAAAGAAATACAACTTAAGTCA 3850

Db 30713 CGTCAAAAAAAGAAAAGAAAAGAAAATGATTAACACCT 30675

RESULT 13

US-09-797-906-3
Sequence 3, Application US/09797906
Patent No. 6329188

GENERAL INFORMATION:

APPLICANT: Zianphe YAN, Karen A. KETCHUM, Valentina DIFRANCESCO, Ellen M. BEASLEY

TITLE OF INVENTION: ISOLATED HUMAN PROTEASE PROTEINS, AND

TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PROTEASE PROTEINS, AND

FILE REFERENCE: CL001151CIP

CURRENT APPLICATION NUMBER: US/09/797,906

CURRENT FILING DATE: 2001-03-05

NUMBER OF SEQ ID NOS: 5

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 3

LENGTH: 84495

TYPE: DNA

ORGANISM: Human

FEATURE:

NAME/KEY: misc_feature

LOCATION: (1)...(84495)

OTHER INFORMATION: n = A,T,C or G

US-09-797-906-3

Query Match 1.7%; Score 66.6; DB 4; Length 84495;
Best Local Similarity 84.3%; Pred. No. 6e-07;
Matches 75; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

Qy 3755 TGGCAGTAGCGGAGATGCACCACTGCACCCCTGCTGGGCGAGAGTGAAGTCTTGT 3814
Db 20107 TTGTAGTGCAGATGCACCACTGCACCTCGGTGGTGCAGAGTGAAGTCTTGT 20166

Qy 3815 CTCATTATACAAAAGAAAAGAAAT 3843

Db 20167 CTCAAAAAAGAAAAGAAAAGAAAT 20195

RESULT 14

US-08-965-048-5/c

Sequence 5, Application US/08965048

Patent No. 6323244

GENERAL INFORMATION:

APPLICANT: Freimer, Nelson

TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE DIAGNOSIS AND

TITLE OF INVENTION: TREATMENT OF NEUROPSYCHIATRIC DISORDERS

FILE REFERENCE: 7853-093

CURRENT APPLICATION NUMBER: US/08/965,048

CURRENT FILING DATE: 1997-11-05

NUMBER OF SEQ ID NOS: 8

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 5

LENGTH: 45716

TYPE: DNA

ORGANISM: Homo sapiens

US-08-965-048-5

Query Match 1.7%; Score 66.4; DB 4; Length 45716;
Best Local Similarity 71.0%; Pred. No. 4.9e-07;
Matches 88; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

Qy 3723 CAGCATATATCTCTAGAAATACAACTAATGCGAGTGGCGAGATGCCAGCACCTGC 3782
Db 3997 CAGGGAATCGCTTAAACCCAGGAGCGAGCTTGCAGTGAAGTGAAGTGAAGTCTTGT 3938

```
QY 3783 ACCCTGCTGGCGGACAGAGTGTCTCTATTACAAAAAGAAAGAAA 3842
    || || || || || || || || || || || || || || || || ||
Db 3937 ACTCCAGCCTGGGCGACAGAGTGTCTCTCAAAATAATAATAAGAAATA 3878
QY 3843 TACA 3846
    | |
Db 3877 AAAA 3874
```

RESULT 15

```
US-08-965-048-6/c
; Sequence 6, Application US/08965048
; Patent No. 6323244
; GENERAL INFORMATION:
; APPLICANT: Chen, Hong
; APPLICANT: Freilmer, Nelson
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE DIAGNOSIS AND
; TITLE OF INVENTION: TREATMENT OF NEUROPSYCHIATRIC DISORDERS
; FILE REFERENCE: 7853-093
; CURRENT APPLICATION NUMBER: US/08/965,048
; CURRENT FILING DATE: 1997-11-05
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
; LENGTH: 45989
; TYPE: DNA
; ORGANISM: Homo sapiens
US-08-965-048-6
```

Query Match

1.7%; Score 66.4; DB 4; Length 45989;

Best Local Similarity 71.0%; Pred.No. 4.9e-07; Mismatches 36; Indels 0; Gaps 0;

```
QY 3723 CAGCATCATATCTCTAGAAATACAACCTAATGGCAGTGTGCGGAGATCGCACACTGC 3782
    || || || || || || || || || || || || || || || || ||
Db 3997 CAGGAGATCGCTTCAACCCAGAGGCGAGGTGCGAGTGCAGCCGAGATCGCACACTGC 3938
QY 3783 ACCCTGCTGGCGGACAGAGTGTCTCTATTACAAAAAGAAAGAAA 3842
    || || || || || || || || || || || || || || || || ||
Db 3937 ACTCCAGCCTGGGCGACAGAGTGTCTCTCAAAATAATAATAAGAAATA 3878
QY 3843 TACA 3846
    | |
Db 3877 AAAA 3874
```

Search completed: November 26, 2002, 21:08:08
Job time : 3852.23 secs

Result No.	Score	Query Match	length	DB	ID	Description
1	3857	100.0	3857	10	US-09-919-497-14	Sequence 14, Appl
2	2857.4	74.1	2900	10	US-09-925-300-670	Sequence 670, Appl
3	1829.4	47.4	3564.1	10	US-09-962-436-306	Sequence 306, Appl
4	1829.4	47.4	3564.1	10	US-09-880-107-2225	Sequence 2225, Appl
5	256.2	6.6	601	10	US-09-867-701-10472	Sequence 10472, A
6	154.6	4.0	3857	10	US-09-919-497-14	Sequence 14, Appl
7	154.6	4.0	3564.1	10	US-09-962-436-306	Sequence 306, Appl
8	154.6	4.0	3564.1	10	US-09-880-107-2225	Sequence 2225, Appl
9	70.2	1.8	9566	10	US-09-764-864-1768	Sequence 1768, Appl
10	70.2	1.8	9566	10	US-09-764-877-3485	Sequence 3485, Appl
11	69.2	1.8	3705	10	US-09-764-877-2510	Sequence 2510, Appl
12	69.2	1.8	145831	10	US-09-969-708-79	Sequence 79, Appl
13	69.2	1.8	145831	10	US-09-954-456-2116	Sequence 2116, Appl
14	69	1.8	24218	9	US-09-860-670-263	Sequence 263, Appl
15	68	1.8	333	10	US-09-764-878-413	Sequence 413, Appl
16	68	1.8	333	10	US-09-764-878-413	Sequence 413, Appl
17	68	1.8	5299	10	US-09-764-878-409	Sequence 409, Appl
18	68	1.8	8210	10	US-09-764-878-406	Sequence 406, Appl
19	67.6	1.8	1400	10	US-09-263-959-295	Sequence 295, Appl

C	20	67.6	1.8	9474	10	US-09-764-878-372	Sequence 372, App
C	21	67.6	1.8	684973	10	US-09-263-959-1	Sequence 1, Appl
C	22	67.4	1.7	279	10	US-09-764-887-507	Sequence 507, App
C	23	67.4	1.7	1668	10	US-09-993-811-13	Sequence 13, Appl
C	24	67.4	1.7	4662	10	US-09-764-869-1902	Sequence 1902, App
C	25	67.4	1.7	4662	10	US-09-764-869-1903	Sequence 1903, App
C	26	67.4	1.7	9192	10	US-09-764-847-1246	Sequence 1246, App
C	27	67.4	1.7	9192	10	US-09-764-877-2218	Sequence 2218, App
C	28	67.4	1.7	9745	10	US-09-764-869-2255	Sequence 2255, App
C	29	67.4	1.7	12149	10	US-09-764-869-2258	Sequence 2258, App
C	30	67.4	1.7	13182	10	US-09-817-199A-3	Sequence 3, Appl
C	31	67.4	1.7	32203	10	US-09-764-869-1849	Sequence 1849, App
C	32	67.4	1.7	368004	10	US-09-949-554-3	Sequence 3, Appl
C	33	67.4	1.7	323	10	US-09-867-701-9630	Sequence 9630, App
C	34	67.2	1.7	23603	9	US-09-860-670-264	Sequence 264, App
C	35	67.2	1.7	23613	9	US-09-860-670-258	Sequence 258, App
C	36	67.2	1.7	24699	10	US-09-764-877-2419	Sequence 2419, App
C	37	67.2	1.7	167343	10	US-09-962-436-281	Sequence 281, App
C	38	67.2	1.7	167343	10	US-09-964-824A-273	Sequence 273, App
C	39	67	1.7	21833	9	US-09-764-877-2275	Sequence 2275, App
C	40	67	1.7	31994	9	US-09-764-904-71	Sequence 71, Appl
C	41	67	1.7	31994	9	US-09-764-860-599	Sequence 599, App
C	42	67	1.7	176373	9	US-10-095-407-17	Sequence 17, Appl
C	43	66.8	1.7	329	10	US-09-764-877-3990	Sequence 3990, App
C	44	66.6	1.7	565	10	US-09-925-301-611	Sequence 611, App
C	45	66.6	1.7	2664	10	US-09-764-887-513	Sequence 513, App

ALIGNMENTS

```

RESULT 1
US-09-919-497-14
: Sequence 14, Application US/09919497
: Patent No. US2002010666A1
: GENERAL INFORMATION:
: APPLICANT: Mutter, George L.
: TITLE OF INVENTION: PROGNOSTIC CLASSIFICATION OF ENDOMETRIAL CANCER
: FILE REFERENCE: B0801/7225
: CURRENT APPLICATION NUMBER: US/09/919,497
: CURRENT FILING DATE: 2001-07-31
: PRIOR APPLICATION NUMBER: US 60/221,735
: PRIOR FILING DATE: 2000-07-31
: NUMBER OF SEQ ID NOS: 100
: SOFTWARE: PatentIn version 3.0
: SEQ ID NO 14
: LENGTH: 3857
: TYPE: DNA
: ORGANISM: Homo sapiens
: US-09-919-497-14

```

```
Query Match      100.0%;  Score 3857;  DB 10;  Length 3857;
```

Best Local Similarity 100.0%; Pred. No. 0;
Matches 3857; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	GGAGAGACCTTCATGAGATCAGGAGGACCGCGCTGCAGCGAATTGCATGGCGCCGCTGCTG	60
Db	1	GGAGAGACCTTCATGAGATCAGGAGGACCGCGCTGCAGCGAATTGCATGGCGCCGCTGCTG	60
QY	61	GAGTAGCAGCGACACGCTGGTGTCTGGAACCTGCTGACACTGACGGCCCTAGTAGTGTGCGCC	120
Db	61	GAGTAGCAGCGACACGCTGGTGTCTGGAACCTGCTGACACTGACGGCCCTAGTAGTGTGCGCC	120
QY	121	CGCGGGCTCGGGCGGAGACCGGCTCTCTACACTTTTCCAGCTGCACCTCCACCCAGCC	180
Db	121	CGCGGGCTCGGGCGGAGACCGGCTCTCTACACTTTTCCAGCTGCACCTCCACCCAGCC	180
QY	181	TGCCCTGGTGTGCTGTGCTCAACACGACCGCGCCGAGGAGGAGTATTTTATCAATCAGCTG	240
Db	181	TGCCCTGGTGTGCTGTGCTCAACACGACCGCGCCGAGGAGGAGTATTTTATCAATCAGCTG	240
QY	241	AAGTAGAGGAGATTGAACACCTCCCTCGCGGTATTAACAAATGAAATCACAGCAACGT	300

Db 241 AAGATGAAGAGTGTGAACACCTCCCTCCGCTGTAAACAATGAATCAACAAGACAGT 300
QY 301 CGCTATGAGGTTTACACAGAGGCGTGTATTTGGCAGACAGTACGATCTTGCTT 360
Db 301 CGCTATGAGGTTTACACAGAGGCGTGTATTTGGCAGACAGTACGATCTTGCTT 360
QY 361 GACTCTTGACATGATGAATACCTTCAGATTTAACTACTGCGATCTTGCTTATAGCC 420
Db 361 GACTCTTGACATGATGAATACCTTCAGATTTAACTACTGCGATCTTGCTTATAGCC 420
QY 421 CACAGATATATGAGCTCTGTCAAGAAGCATTTCTTGCGCTCTTGCGCCAGAAAAC 480
Db 421 CACAGATATATGAGCTCTGTCAAGAAGCATTTCTTGCGCTCTTGCGCCAGAAAAC 480
QY 481 AAACGTGTTTTTAAAGCTTTCACAGACAAATGCTTGCTCTTGATACGTTTTTGT 540
Db 481 AAACGTGTTTTTAAAGCTTTCACAGACAAATGCTTGCTCTTGATACGTTTTTGT 540
QY 541 CATGTGGAAGAGTGTATGAGAAATCTTTTGTGAGAAACCTGTATCTGCGCCAGAAAAC 600
Db 541 CATGTGGAAGAGTGTATGAGAAATCTTTTGTGAGAAACCTGTATCTGCGCCAGAAAAC 600
QY 601 CATGTGAGCAATACATCTTTTGAAGACAGCAAACTGAATGTAGAAATCCATGTT 660
Db 601 CATGTGAGCAATACATCTTTTGAAGACAGCAAACTGAATGTAGAAATCCATGTT 660
QY 661 TCTATGACACCTTACATGCTTGTCTATACAGACTGTATCTACTGACATTTTAAATGCAATG 720
Db 661 TCTATGACACCTTACATGCTTGTCTATACAGACTGTATCTACTGACATTTTAAATGCAATG 720
QY 721 CTAAAGGAAGTAAATGCGCTTAACCCATGCTTGAAGTGAAGATTTATCTTTAGAAAT 780
Db 721 CTAAAGGAAGTAAATGCGCTTAACCCATGCTTGAAGTGAAGATTTATCTTTAGAAAT 780
QY 781 GCTATTGGAAGACCTTTTGACAGACATCCGCCATTTATCTGATCTTTTGCGACAG 840
Db 781 GCTATTGGAAGACCTTTTGACAGACATCCGCCATTTATCTGATCTTTTGCGACAG 840
QY 841 CTGTGAGCCAAAGACTAAATCTTATGTTCAAGATTTGAAGATTTAGCAACTTGTGCGAG 900
Db 841 CTGTGAGCCAAAGACTAAATCTTATGTTCAAGATTTGAAGATTTAGCAACTTGTGCGAG 900
QY 901 TATCTCTCAGTATGATGTTGTACATTTCTTAATCTCTGGAATCTCGAAGCAAGC 960
Db 901 TATCTCTCAGTATGATGTTGTGTACATTTCTTAATCTCTGGAATCTCGAAGCAAGC 960
QY 961 GAAAAAGCTTTTGTGAGATTCAGATTCAGTGTCTTCTTGACTCAGCACTCGATGTTT 1020
Db 961 GAAAAAGCTTTTGTGAGATTCAGATTCAGTGTGTCTTCTTGACTCAGCACTCGATGTTT 1020
QY 1021 ATTAATGCTGAGCAAGGTTTATCATCTTCCAGATGCCAAATGAGTAAAGAAAA 1080
Db 1021 ATTAATGCTGAGCAAGGTTTATCATCTTCCAGATGCCAAATGAGTAAAGAAAA 1080
QY 1081 ATATCTGAAAAAATGAAAAATTAAGAAGGGAAGAAAAAGAAAGAACTGCTCTAGAA 1140
Db 1081 ATATCTGAAAAAATGAAAAATTAAGAAGGGAAGAAAAAGAAAGAACTGCTCTAGAA 1140
QY 1141 AGCAACCCAAAGTGGAGGAGCTGATGAATTTAAAGAAATTAAGCAGAAATTAAG 1200
Db 1141 AGCAACCCAAAGTGGAGGAGCTGATGAATTTAAAGAAATTAAGCAGAAATTAAG 1200
QY 1201 GAGAGTGAAGCTTGTGTGTCAGGTCAAGTACTGATTTGTGTCAGAGTGAAGCCAGACA 1260
Db 1201 GAGAGTGAAGCTTGTGTGTCAGGTCAAGTACTGATTTGTGTCAGAGTGAAGCCAGACA 1260
QY 1261 TGTTCCTCAGCTGAGAGCTATATCACTTTTGAGCGGAGGCTTTTATTGAGGCTCTAC 1320
Db 1261 TGTTCCTCAGCTGAGAGCTATATCACTTTTGAGCGGAGGCTTTTATTGAGGCTCTAC 1320
QY 1321 AGCAAAACCTTTGAGAAGGTAGCAAAAGCGGAAGATGCGATGTAATTTAGCAAGGAA 1380
Db 1321 AGCAAAACCTTTGAGAAGGTAGCAAAAGCGTGAAGAGTGTGATGAATTTAGCAAGGAA 1380
QY 1381 GACAGTTCAAGAGATTTAGGAATCTCAAAAAGACCTTAAGACCCCAAAAAGAA 1440
Db 1381 GACAGTTCAAGAGATTTAGGAATCTCAAAAAGACCTTAAGACCCCAAAAAGAA 1440
QY 1441 CGGGCTTCTACCAAGAAAGAACCCCTCAAAAAGAAAAAGGAAATCTGACCTTAAC 1500
Db 1441 CGGGCTTCTACCAAGAAAGAACCCCTCAAAAAGAAAAAGGAAATCTGACCTTAAC 1500
QY 1501 ATGCTGAAAAACCTGAAAGTGAAGGAGGAGATGTGAGCAATGATGCTCA 1560
Db 1501 ATGCTGAAAAACCTGAAAGTGAAGGAGGAGATGTGAGCAATGATGCTCA 1560
QY 1561 GAAATGAGAGTGAAGCCAGAAAGCTCCGGAAGAAATTAAGCATGMAAATTTGATGA 1620
Db 1561 GAAATGAGAGTGAAGCCAGAAAGCTCCGGAAGAAATTAAGCATGMAAATTTGATGA 1620
QY 1621 AATTTGTATCGATGCTGCTTTGGAATCTGAAAGAACCCCTCAATCATCCATCGG 1680
Db 1621 AATTTGTATCGATGCTGCTTTGGAATCTGAAAGAACCCCTCAATCATCCATCGG 1680
QY 1681 CTTCGTGTTGACGAGCCCTATGCTGTGACAAAGGGTACTGATGATGAGCCAGAA 1740
Db 1681 CTTCGTGTTGACGAGCCCTATGCTGTGACAAAGGGTACTGATGATGAGCCAGAA 1740
QY 1741 TACGTGTTCTTTATGACGACAGAGTAACTTGTTCGACACTTGAATTTTACAGGGG 1800
Db 1741 TACGTGTTCTTTATGACGACAGAGTAACTTGTTCGACACTTGAATTTTACAGGGG 1800
QY 1801 AGTAGGCTTGGAAACCTGTGAGGGTTTACTTCTTATATACGGAGGTTTCACTGAGAA 1860
Db 1801 AGTAGGCTTGGAAACCTGTGAGGGTTTACTTCTTATATACGGAGGTTTCACTGAGAA 1860
QY 1861 CAACGCTATCTCACTGCTTTTGGGGAAGAAAGAAAGCTTTTGAAMAACCTAAGGGA 1920
Db 1861 CAACGCTATCTCACTGCTTTTGGGGAAGAAAGAAAGCTTTTGAAMAACCTAAGGGA 1920
QY 1921 AAAGCAAGATGTTTCCCTGAAGAAAGAAAGGACAGATGAACAACCTTAGACCTA 1980
Db 1921 AAAGCAAGATGTTTCCCTGAAGAAAGAAAGGACAGATGAACAACCTTAGACCTA 1980
QY 1981 GTTAGAGGACAGACATCTGAGATGTTTCCAGTGAACCTGGAACCCGCTGGACGAA 2040
Db 1981 GTTAGAGGACAGACATCTGAGATGTTTCCAGTGAACCTGGAACCCGCTGGACGAA 2040
QY 2041 CAGAAATGTACACAGCAAAAGCATAGTTGTGATATGCGAATTTGGAAGTGAAGCTTCA 2100
Db 2041 CAGAAATGTACACAGCAAAAGCATAGTTGTGATATGCGAATTTGGAAGTGAAGCTTCA 2100
QY 2101 TCTCTGATCATGCTGGGGCATTTGACATTTGAACCCGTACTTTATAGTTGAGATTTAC 2160
Db 2101 TCTCTGATCATGCTGGGGCATTTGACATTTGAACCCGTACTTTATAGTTGAGATTTAC 2160
QY 2161 ATCCCTGATCCAGAAATGTGCGGTGAGGCGCAAGAGTATGAGTATTAATGCGCTTTTA 2220
Db 2161 ATCCCTGATCCAGAAATGTGCGGTGAGGCGCAAGAGTATGAGTATTAATGCGCTTTTA 2220
QY 2221 AATTAAGCGCGCTCTACAGCCAGTGCATCTCCATCTCCCGCTACTTCAAGCGTCCCGT 2280
Db 2221 AATTAAGCGCGCTCTACAGCCAGTGCATCTCCATCTCCCGCTACTTCAAGCGTCCCGT 2280
QY 2281 CTTCGATTTGAGTTGACCTTACCAAGGCTTTCTCTCAGCTTCCGAGAGTGCCTTGT 2340
Db 2281 CTTCGATTTGAGTTGACCTTACCAAGGCTTTCTCTCAGCTTCCGAGAGTGCCTTGT 2340
QY 2341 CAGGAGATCTCCAGCAATACATTAATTTCCAAACTACCTCTTCAACTTCACTTCCCG 2400
Db 2341 CAGGAGATCTCCAGCAATACATTAATTTCCAAACTACCTCTTCAACTTCACTTCCCG 2400
QY 2401 AGACTACGATTTCTGTGGCCCTCTCTCATGCAAGCGGAGTGTGTTGAGAGAGCTG 2460
Db 2401 AGACTACGATTTCTGTGGCCCTCTCTCATGCAAGCGGAGTGTGTTGAGAGAGCTG 2460

```

Oy 2461 AAACAAAGCAGCAGCAGCTGATGCGGCGACAGCAGCTGGCCATTACAGAGATTCTGAA 2520
    |||
Db 2461 AAACAAAGCAGCAGCAGCTGATGCGGCGACAGCAGCTGGCCATTACAGAGATTCTGAA 2520
Oy 2521 ACCCTCCGAGTGCAGAGAGATATATCTGCTCCCAAGACTTCTGTATAAATGCCA 2580
    |||
Db 2521 ACCCTCCGAGTGCAGAGAGATATATCTGCTCCCAAGACTTCTGTATAAATGCCA 2580
Oy 2581 GGGGGAATGCAAAAACGCGCTCTGTATGACACACAGTAAAGACATGCGAGAAATTA 2640
    |||
Db 2581 GGGGGAATGCAAAAACGCGCTCTGTATGACACACAGTAAAGACATGCGAGAAATTA 2640
Oy 2641 GCAGCCCTGTCACAAGAGAGCTCAGAGTATTCGGGGAAATGCTGCAAAATGCCAAACAG 2700
    |||
Db 2641 GCAGCCCTGTCACAAGAGAGCTCAGAGTATTCGGGGAAATGCTGCAAAATGCCAAACAG 2700
Oy 2701 CTTTATGATTTATTCACACCTCTTTTGCAGAGTCGATCAAAAGGAAAGGAAAAAG 2760
    |||
Db 2701 CTTTATGATTTATTCACACCTCTTTTGCAGAGTCGATCAAAAGGAAAGGAAAAAG 2760
Oy 2761 TGAACAGTATGCTGTTTCTTATCCATGCTGATCTTTCAGGGGCTCCTTGCCAGA 2820
    |||
Db 2761 TGAACAGTATGCTGTTTCTTATCCATGCTGATCTTTCAGGGGCTCCTTGCCAGA 2820
Oy 2821 CATCATAGGTCATTTATTTATTTGTTGCTATTTTCATTTCTTTTCCAAATGCTTTAATG 2880
    |||
Db 2821 CATCATAGGTCATTTATTTATTTGTTGCTATTTTCATTTCTTTTCCAAATGCTTTAATG 2880
Oy 2881 ATTGACGGTGCAGCAGAGAGCAGATTCCTCTGACACTGCAATTAGGCATCACTTG 2940
    |||
Db 2881 ATTGACGGTGCAGCAGAGAGCAGATTCCTCTGACACTGCAATTAGGCATCACTTG 2940
Oy 2941 AACTGGCGTGTGCTGCTCTTTTCTCCCTGACGCTGATGCGGGGCTTGAGCATGTT 3000
    |||
Db 2941 AACTGGCGTGTGCTGCTCTTTTCTCCCTGACGCTGATGCGGGGCTTGAGCATGTT 3000
Oy 3001 CTTTAAATGAGGTTTGTGACGAGTACAGTAAGTTCCTACAAGTATTCAGAAAGTAG 3060
    |||
Db 3001 CTTTAAATGAGGTTTGTGACGAGTACAGTAAGTTCCTACAAGTATTCAGAAAGTAG 3060
Oy 3061 AAACCTTACCTGATCTCTACAGATCTCATTTGAAAAGAAATATGCTAACCTGCGCATGA 3120
    |||
Db 3061 AAACCTTACCTGATCTCTACAGATCTCATTTGAAAAGAAATATGCTAACCTGCGCATGA 3120
Oy 3121 CGGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 3180
    |||
Db 3121 CGGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 3180
Oy 3181 TCTAGTGCACATTAATAACAGCTTTAATGCACTTATACCCATGCTGCTGCTGCCA 3240
    |||
Db 3181 TCTAGTGCACATTAATAACAGCTTTAATGCACTTATACCCATGCTGCTGCTGCCA 3240
Oy 3241 AATCGGTCTTGTGCTGTGCTGCTGCTGAGCCTTGAACGTGATGTTGTGTGGAATCA 3300
    |||
Db 3241 AATCGGTCTTGTGCTGTGCTGCTGCTGAGCCTTGAACGTGATGTTGTGTGGAATCA 3300
Oy 3301 TGTGTGACCCCTTGTCTACAAAGAGGCTTGTGGAACAGTGAAGAAACATCTCTTGG 3360
    |||
Db 3301 TGTGTGACCCCTTGTCTACAAAGAGGCTTGTGGAACAGTGAAGAAACATCTCTTGG 3360
Oy 3361 CCATTCCTGACACAGTGTCTCTACACATTTTCTTCAGTCCATGCTGCTGCTGCTG 3420
    |||
Db 3361 CCATTCCTGACACAGTGTCTCTACACATTTTCTTCAGTCCATGCTGCTGCTGCTG 3420
Oy 3421 TCTAAGGAAATTTTCTAGAGGCTTCTCTACTATTCAGAGACTGCTCTCAAAACG 3480
    |||
Db 3421 TCTAAGGAAATTTTCTAGAGGCTTCTCTACTATTCAGAGACTGCTCTCAAAACG 3480
Oy 3481 GTTGACTAGTCTTCTAATGACCCCTAATATGTAGCATATATATTAATTCTTGTGCA 3540
    |||
Db 3481 GTTGACTAGTCTTCTAATGACCCCTAATATGTAGCATATATATTAATTCTTGTGCA 3540
Oy 3541 AATTAGTATTTTAAAGCAAAATGAATTAACCTGTTGCAAAAAGTTAATGATGAAGAGCT 3600
    |||

```

```

Db 3541 AATTAGTATTTTAAAGCAAAATGAATTAACCTGTTTGCAAAAGTTAATGATGAAGAGCT 3600
    |||
Oy 3601 CTTAGAAATTCGAATTTTGGACATATTCAGTCTCTCTATATACAGAGATCCCTAGTCCA 3660
    |||
Db 3601 CTTAGAAATTCGAATTTTGGACATATTCAGTCTCTCTATATACAGAGATCCCTAGTCCA 3660
Oy 3661 GCTGGCTAGTTCAGAGTATTTTTCAGACTTCTCTGCTCTCAGCTCTTATCTTAAGAC 3720
    |||
Db 3661 GCTGGCTAGTTCAGAGTATTTTTCAGACTTCTCTGCTCTCAGCTCTTATCTTAAGAC 3720
Oy 3721 ACCAGCATATATCTCTAGAAATACAACTAATTTGGCAGTGAAGCCGAGATGCAACACT 3780
    |||
Db 3721 ACCAGCATATATCTCTAGAAATACAACTAATTTGGCAGTGAAGCCGAGATGCAACACT 3780
Oy 3781 GCACCCCTGCTGGGGGAGCAGAGTGAAGCTTGTCTATTAACAAAAAGAAAAAGA 3840
    |||
Db 3781 GCACCCCTGCTGGGGGAGCAGAGTGAAGCTTGTCTATTAACAAAAAGAAAAAGA 3840
Oy 3841 AATACAACTAAGCTCA 3857
    |||
Db 3841 AATACAACTAAGCTCA 3857
    |||

```

```

RESULT 2
US-09-925-300-670
; Sequence 670, Application US/09925300
; Patent No. US20020151681A1
; GENERAL INFORMATION:
; APPLICANT: Craig Rosen,
; APPLICANT: Steve Ruben
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA101
; CURRENT APPLICATION NUMBER: US/09/925,300
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05988
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; NUMBER OF SEQ ID NOS: 1890
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 670
; LENGTH: 2900
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (2418)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-925-300-670

```

```

Query Match 74.1%; Score 2857.4; DB 10; Length 2900;
Best Local Similarity 99.6%; Pred. No. 0;
Matches 2869; Conservative 4; Mismatches 6; Indels 1; Gaps 1;
Oy 25 GAGCGGCTCGAGCGGATTCATGCGGCGGCTGAGTACAGCAGAGTGGGCTG 84
    |||
Db 15 GCGCGGCTCGAGCGGATTCATGCGGCGGCTGAGTACAGCAGAGTGGGCTG 73
Oy 85 GAATGCTCGACAGTGAAGGCTAGTGTGCGGCGGCGGCTGCGCGGAGCGGCTC 144
    |||
Db 74 GAATGCTCGACAGTGAAGGCTAGTGTGCGGCGGCGGCTGCGCGGAGCGGCTC 133
Oy 145 CTTTACCACTTCTCAGTGCACGTGCAACCCAGGCTGCTGTGCTGTCAACAG 204
    |||
Db 134 CTTTACCACTTCTCAGTGCACGTGCAACCCAGGCTGCTGTGCTGTCAACAG 193
Oy 205 CAGCGGCGGAGAGAGATTTTATCAATCAGTGAAGATGAAGGACTTGAACACTC 264
    |||
Db 194 CAGCGGCGGAGAGAGATTTTATCAATCAGTGAAGATGAAGGACTTGAACACTC 253
Oy 265 CCTGCGGCTGAACAAATGAATCAACAGCACTGCTATGAAGTTTACACAGAGT 324
    |||

```


D	254	CTGCGCCGTGAACAAATGAAATCACAGCAACAGTCGTATGAAATTTACACCAAGT	313
Q	325	GGTGTATATTTGGCAAAATAGATACTTGTGGTTGACTTCTTGATGATAGAAATCCT	384
D	314	GGGTATATTTGGCAAGAGATGATCTTGTGGTTGACTTCTTGATGATAGAAATCCT	373
Q	385	TCAGATTTAAATTAAGTCGATCTTGTGTATAGAGCCACAGAAATTAATGATCTTGCA	444
D	374	TCAGATTTAAATTAAGTCGATCTTGTGTATAGAGCCACAGAAATTAATGATCTTGCA	433
Q	445	GAAGCATTCATCTTGGCCCTCTTCCGACAGAAAACAAGCTGTTTATTAAGCTTTC	504
D	434	GAAGCATTCATCTTGGCCCTCTTCCGACAGAAAACAAGCTGTTTATTAAGCTTTC	493
Q	505	ACAGACATGCTGTGGCTTGTGATGATGATGATGATGATGATGATGATGATGATGAT	564
D	494	ACAGACATGCTGTGGCTTGTGATGATGATGATGATGATGATGATGATGATGATGAT	553
Q	565	CTTTTGTGAGAAACTGTATCTGTGGCCAAAGTTCCATGTAGCAGTAACTATTTT	624
D	554	CTTTTGTGAGAAACTGTATCTGTGGCCAAAGTTCCATGTAGCAGTAACTATTTT	613
Q	625	GAACAGACAAACCTGAAGTTGAGAAATCCATGTTCTATGACACCTACCAGCTTGT	684
D	614	GAACAGACAAACCTGAAGTTGAGAAATCCATGTTCTATGACACCTACCAGCTTGT	673
Q	685	ATACAGCTGCTATACGACATTTTAAATGATGATGATGATGATGATGATGATGATGAT	744
D	674	ATACAGCTGCTATACGACATTTTAAATGATGATGATGATGATGATGATGATGATGAT	733
Q	745	CCATCGCTTGAAGTGAAGATTTATCTTGAAGAAATGCTATTTGAGAAACCTTTGAC	804
D	734	CCATCGCTTGAAGTGAAGATTTATCTTGAAGAAATGCTATTTGAGAAACCTTTGAC	793
Q	805	ACAATCCGCAATATCTGATCCTTTGTGCACACGCTTGAGCCCAAGACTAACTCTTA	864
D	794	ACAATCCGCAATATCTGATCCTTTGTGCACACGCTTGAGCCCAAGACTAACTCTTA	853
Q	865	GTTGAGATTTGAAGATATACGAACCTTGTGATGATGATGATGATGATGATGATGATGAT	924
D	854	GTTGAGATTTGAAGATATACGAACCTTGTGATGATGATGATGATGATGATGATGATGAT	913
Q	925	ACATCTTATATCTTGTGATCCTGTGAGACCAACGAAAGCTTTTGGTCAGAAATTA	984
D	914	ACATCTTATATCTTGTGATCCTGTGAGACCAACGAAAGCTTTTGGTCAGAAATTA	973
Q	985	GGTGGCTGTTTCTGACTCCAGCAGCTGATTTTAAATGCTGAGCAAGGCTTAT	1044
D	974	GGTGGCTGTTTCTGACTCCAGCAGCTGATTTTAAATGCTGAGCAAGGCTTAT	1033
Q	1045	CATCTTCAGATGCAAAATGAGTAAAGAAATATCTGAAATTAATGAAATTA	1104
D	1034	CATCTTCAGATGCAAAATGAGTAAAGAAATATCTGAAATTAATGAAATTA	1093
Q	1105	GAAGGGGAGAAACAAAGAAAGAACTGCTAGAAAGCAACCCAAAGTGGAGGCACTG	1164
D	1094	GAAGGGGAGAAACAAAGAAAGAACTGCTAGAAAGCAACCCAAAGTGGAGGCACTG	1153
Q	1165	ACTGAAGTATTAAGAAATTAAGGAGCAAAATTAAGGAGTGAAGCTCTTGGTGCA	1224
D	1154	ACTGAAGTATTAAGAAATTAAGGAGCAAAATTAAGGAGTGAAGCTCTTGGTGCA	1213
Q	1225	GGTCAAGTACTGATTTGTGAGATGAGCAAGATGTTCCAGCTGAGAGACTATATC	1284
D	1214	GGTCAAGTACTGATTTGTGAGATGAGCAAGATGTTCCAGCTGAGAGACTATATC	1273
Q	1285	ACTCTGGAGCGAGGCTTCTTATTTAGGCTCTACAGAAAACCTTTGAGAGGATAGC	1344
D	1274	ACTCTGGAGCGAGGCTTCTTATTTAGGCTCTACAGAAAACCTTTGAGAGGATAGC	1333
Q	1345	AAAGCTGAGAGAGTGTGATGAATTTAGAGAGAGACAGTTTCAAGAGATTTAGGAA	1404
D	1334	AAAGCTGAGAGAGTGTGATGAATTTAGAGAGAGACAGTTTCAAGAGATTTAGGAA	1393

Q	1405	TTCTCAAAAAGACCTAAAGACCCCAAAACAAAGAGGCTTCTATCAAAAGAAAGAC	1464
D	1394	TTCTCAAAAAGACCTAAAGACCCCAAAACAAAGAGGCTTCTATCAAAAGAAAGAC	1453
Q	1465	CTCAAAAAGAAACGGAAGTTGACCTTAATCTCAATGTAGAGAAACCTGGAAGACTG	1524
D	1454	CTCAAAAAGAAACGGAAGTTGACCTTAATCTCAATGTAGAGAAACCTGGAAGACTG	1513
Q	1525	GAAGAGAGAGAGATGTGAGAGAGATATGCTGAGAAATTAAGCAATAGCCGGAAGC	1584
D	1514	GAAGAGAGAGAGATGTGAGAGAGATATGCTGAGAAATTAAGCAATAGCCGGAAGC	1573
Q	1585	TGCCGGAAGAAATTAAGCAATGATGATTAATTTGATATGATGATGATGATGATGATGAT	1644
D	1574	TGCCGGAAGAAATTAAGCAATGATGATTAATTTGATATGATGATGATGATGATGATGAT	1633
Q	1645	GGAATCCTGAAAGAACCCCTCATATCATCCGCTTCTGGGTTGCAGCGACCCCTAT	1704
D	1634	GGAATCCTGAAAGAACCCCTCATATCATCCGCTTCTGGGTTGCAGCGACCCCTAT	1693
Q	1705	GCTTGACAAAGGCTACTACATGAAGTGAGCCAAAGATACGTGTTCTTTATGACGAG	1764
D	1694	GCTTGACAAAGGCTACTACATGAAGTGAGCCAAAGATACGTGTTCTTTATGACGAG	1753
Q	1765	CTAACCTTGTGGGCGAGCTTGAATTTACAGGCGAGTGGCTGGGAAACCTGTAGG	1824
D	1754	CTAACCTTGTGTGGCGAGCTTGAATTTACAGGCGAGTGGCTGGGAAACCTGTAGG	1813
Q	1825	GTTTACTTCTTATATACGAGGTTTCACTGAGAGAACGCTATCTCACTGCTTTCGG	1884
D	1814	GTTTACTTCTTATATACGAGGTTTCACTGAGAGAACGCTATCTCACTGCTTTCGG	1873
Q	1885	AAAGAAAAGAGCTTTTGAAGAACTCATAGGAAAGCAAGCATGTTTGTCCGAA	1944
D	1874	AAAGAAAAGAGCTTTTGAAGAACTCATAGGAAAGCAAGCATGTTTGTCCGAA	1933
Q	1945	GAAAGAGAGGAGATGAAACAACTTGAAGTAAAGAGCAAGCATGTTTGTCCGAA	2004
D	1934	GAAAGAGAGGAGATGAAACAACTTGAAGTAAAGAGCAAGCATGTTTGTCCGAA	1993
Q	2005	GTTTCCATGACACTCGGAAGCGGTGGCCAGAGCAAGATGTTTACAGCAAGCAT	2064
D	1994	GTTTCCATGACACTCGGAAGCGGTGGCCAGAGCAAGATGTTTACAGCAAGCAT	2053
Q	2065	GTTGTGATATGCTGATTTTCAAGTGAAGCTTCTGATCTCTGCTGCGGCAAT	2124
D	2054	GTTGTGATATGCTGATTTTCAAGTGAAGCTTCTGATCTCTGATCTCTGCGGCAAT	2113
Q	2125	GACATTAACCCGATGATTTAGAGGTTGAGATTAATCTGATCTCTGATCTCTGATCTCTGAT	2184
D	2114	GACATTAACCCGATGATTTAGAGGTTGAGATTAATCTGATCTCTGATCTCTGATCTCTGAT	2173
Q	2185	GAGCGAAGATATCACTGATTTAATCGCTCTTAAATTAAGCGCCCTCTACAGCAG	2244
D	2174	GAGCGAAGATATCACTGATTTAATCGCTCTTAAATTAAGCGCCCTCTACAGCAG	2233
Q	2245	TGCATCTCATATGTCGCCCTACTACAGCTGCTGCTTCTGATTTGATTTGACCTAGC	2304
D	2234	TGCATCTCATATGTCGCCCTACTACAGCTGCTGCTTCTGATTTGATTTGACCTAGC	2293
Q	2305	AAGCCTTCTCTCACTTCCCGAGGCTTGTGAGAGATCTGACAGATGACAT	2364
D	2294	AAGCCTTCTCTCACTTCCCGAGGCTTGTGAGAGATCTGACAGATGACAT	2353
Q	2365	AGTTCCAAACTCACTTCTTACACTTCACTTCCAGACATGAGATTTCTGTGCTCC	2424
D	2354	AGTTCCAAACTCACTTCTTACACTTCACTTCCAGACATGAGATTTCTGTGCTCC	2413
Q	2425	TTCTCTCATGCAAGCGGAGTTGTTGAGAGCTGAAACAAAGCAAGCCAGCCTGAT	2484
D	2414	TTCTCTCATGCAAGCGGAGTTGTTGAGAGCTGAAACAAAGCAAGCCAGCCTGAT	2473

Oy 2485 GCGGACAGACACTGGCCATTACAGCAGATTCTGAAACCCCTTCCGAGTCAGAGAATAT 2544
|||||
Db 2474 GCGGAGACAGACACTGGCCATTACAGCAGATTCTGAAACCCCTTCCGAGTCAGAGAATAT 2533
Oy 2545 AATCTGCTGCCCAAGACTCTTGTAAATAATGCCAGGGGTGAATGCCAAAACTGCCGC 2604
|||||
Db 2534 AATCTGCTGCCCAAGACTCTTGTAAATAATGCCAGGGGTGAATGCCAAAACTGCCGC 2593
Oy 2605 TCCGTGATGACACACGATTAAGACATCGCAATATTAGCAGCCCTGTACAGACAGCTC 2664
|||||
Db 2594 TCCGTGATGACACACGATTAAGACATCGCAATATTAGCAGCCCTGTACAGACAGCTC 2653
Oy 2665 ACAGATATTCTGGGAGATGCTGCAAAATGCCAAACAGCTTTATGATTTATTCACACCTC 2724
|||||
Db 2654 ACAGATATTCTGGGAGATGCTGCAAAATGCCAAACAGCTTTATGATTTATTCACACCTC 2713
Oy 2725 TTTGAGAGATGCTGCAAAAGGAAAGGAAAGGAAAGGAAAGGAAAGGAAAGGAAAGGAAAG 2784
|||||
Db 2714 TTTGAGAGATGCTGCAAAAGGAAAGGAAAGGAAAGGAAAGGAAAGGAAAGGAAAGGAAAG 2773
Oy 2785 TCCCATGCTGCTACTTTTCAGAGGCTCTTGCAGACATCATATGATTTATTTATTTAT 2844
|||||
Db 2774 TCCCATGCTGCTACTTTTCAGAGGCTCTTGCAGACATCATATGATTTATTTATTTAT 2833
Oy 2845 GGTTCGCTATTTTCATTTCTTTCCTCAATGCTCTTAATGATTTGACGCTGACAGACGAG 2904
|||||
Db 2834 GGTTCGCTATTTTCATTTCTTTCCTCAATGCTCTTAATGATTTGACGCTGACAGACGAG 2893

RESULT 3

US-09-962-436-306
; Sequence 306, Application US/09962436
; Patent No. US20020081301A1
; GENERAL INFORMATION:
; APPLICANT: Soppet, Daniel
; TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using Signatu
; FILE REFERENCE: 689290-75
; CURRENT APPLICATION NUMBER: US/09/962, 436
; PRIOR FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: US/60/235, 082
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US/60/234, 924
; NUMBER OF SEQ ID NOS: 568
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 306
; LENGTH: 35641
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-962-436-306

Query Match 47.4%; Score 1829.4; DB 10; Length 35641;
Best Local Similarity 99.4%; Pred. No. 0;
Matches 1836; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

Oy 2011 ACTGACACTCGGAAAGCCGCTGCGCCAGAGACAGAAATGTTAGACAGCAAAAGCATAGTTG 2070
|||||
Db 29522 ACTTACTTTTCTCTGTAGTGGCCAGAGACAGAAATGTTAGACAGCAAAAGCATAGTTG 29581
Oy 2071 GATATGCTGATTTTCAAGTGAAGCTTCCATCTCTGATCCATCGTGGGGCATTTGACATT 2130
|||||
Db 29582 GATATGCTGATTTTCAAGTGAAGCTTCCATCTCTGATCCATCGTGGGGCATTTGACATT 29641
Oy 2131 GAACCCGAGCTTAGAGTGGAGATTACATCTCAGTCCAGAAATGGGTGGAGGCG 2190
|||||
Db 29642 GAACCCGAGCTTAGAGTGGAGATTACATCTCAGTCCAGAAATGGGTGGAGGCG 29701
Oy 2191 AAGAGTACAGTGAATTAATCGGCTCTTTAAATAACGGCCGCTCTACAGCAGTGCATC 2250
|||||
Db 29702 AAGAGTACAGTGAATTAATCGGCTCTTTAAATAACGGCCGCTCTACAGCAGTGCATC 29761
Oy 2251 TCCATGTCGCCGCTACTACAGCGTCGCCGCTCTGATTTGAGTTGACCCTAGCAAGCTT 2310

|||||
Db 29762 TCCATGTCGCCGCTACTACAGCGTCCCGTCTTCTGATTTGATTTGACCTTAGCAAGCTT 29821
Oy 2311 TTCTCTCTACTTCCCGAGAGTCCCTGTTTTCAGAGATCTCCAGCAATGCATTAAGTTCC 2370
|||||
Db 29822 TTCTCTCTACTTCCCGAGAGTCCCTGTTTTCAGAGATCTCCAGCAATGCATTAAGTTCC 29881
Oy 2371 AAACATCTCTTCTTACACTTACTTCCCGAGACTACAGGATTTCTGTTGGCCCTCTCT 2430
|||||
Db 29882 AAACATCTCTTCTTACACTTACTTCCCGAGACTACAGGATTTCTGTTGGCCCTCTCT 29941
Oy 2431 CATGCAAGCGGAGAGTTGTTGAGAGCTGAAACAAGCAACCCACAGCTGATGCGGGG 2490
|||||
Db 29942 CATGCAAGCGGAGAGTTGTTGAGAGCTGAAACAAGCAACCCACAGCTGATGCGGGG 30001
Oy 2491 ACAGCACTGGCCATTACAGCAGATTTCTGAACCCCTTCCGAGTCAAGAAATATATCT 2550
|||||
Db 30002 ACAGCACTGGCCATTACAGCAGATTTCTGAACCCCTTCCGAGTCAAGAAATATATCT 30061
Oy 2551 GGTCCCAAGACTCTTGTAAATAATGCCAGGGGTGAATGCCAAACCTGCCGCTCTTG 2610
|||||
Db 30062 GGTCCCAAGACTCTTGTAAATAATGCCAGGGGTGAATGCCAAACCTGCCGCTCTTG 30121
Oy 2611 ATGCACACGTTTAAAGACATGCGCAGAAATTAAGCAGCCCTGTCAAGACAGAGCTCAGAGT 2670
|||||
Db 30122 ATGCACACGTTTAAAGACATGCGCAGAAATTAAGCAGCCCTGTCAAGACAGAGCTCAGAGT 30181
Oy 2671 ATTCTGGGAAATGCTGCAAATGCCAAACGCTTTATGATTTTATTCACACCTCTTTTGA 2730
|||||
Db 30182 ATTCTGGGAAATGCTGCAAATGCCAAACGCTTTATGATTTTATTCACACCTCTTTTGA 30241
Oy 2731 GAAGTCGATCAAAAGGAAAGGAAAGGAAAGGAAAGGAAAGGAAAGGAAAGGAAAGGAAAG 2790
|||||
Db 30242 GAAGTCGATCAAAAGGAAAGGAAAGGAAAGGAAAGGAAAGGAAAGGAAAGGAAAGGAAAG 30301
Oy 2791 GCCTTACTTTTCACGCGGCTCTTGCACAGACATCATATGATTTATTTATTTGTTTG 2850
|||||
Db 30302 GCCTTACTTTTCACGCGGCTCTTGCACAGACATCATATGATTTATTTATTTGTTTG 30361
Oy 2851 CTATTTCACTCTTTTCCCAATGCTCTTAATGATTTGACGCTGACACGAAAGCCAGATTC 2910
|||||
Db 30362 CTATTTCACTCTTTTCCCAATGCTCTTAATGATTTGACGCTGACACGAAAGCCAGATTC 30421
Oy 2911 TCTCTGAACCTGCAAGTTAGCATCATGTAACCTGCGTGCCTCTTCTCTCC 2970
|||||
Db 30422 TCTCTGAACCTGCAAGTTAGCATCATGTAACCTGCGTGCCTCTTCTCTCC 30481
Oy 2971 TGCACCGTCTATGCCGGCTTAGCATGTTCTTTTAAATGAGTTTCTCAGATCAGGT 3030
|||||
Db 30482 TGCACCGTCTATGCCGGCTTAGCATGTTCTTTTAAATGAGTTTCTCAGATCAGGT 30541
Oy 3031 AAAGTTCTTACAGAGATTTACAGAGGTTAGAAACTTTACTCTATCTCTAACATCTCATTT 3090
|||||
Db 30542 AAAGTTCTTACAGAGATTTACAGAGGTTAGAAACTTTACTCTATCTCTAACATCTCATTT 30601
Oy 3091 TAGAAGGAATATGCTTAAGCCCTGCGATGACGCTCAGGAGGAGGAAAGAGCAGGACAAA 3150
|||||
Db 30602 TAGAAGGAATATGCTTAAGCCCTGCGATGACGCTCAGGAGGAGGAAAGAGCAGGACAAA 30661
Oy 3151 GAAAGCTACATTTTAAACAGTCTTGTATCTAGTGCACATTAATAACAGTCTTAATT 3210
|||||
Db 30662 GAAAGCTACATTTTAAACAGTCTTGTATCTAGTGCACATTAATAACAGTCTTAATT 30721
Oy 3211 GCACCTTATACCATGCTCGCTGCGCTCCCAATCTGGCTTGGCTGTGTGCTCTGGA 3270
|||||
Db 30722 GCACCTTATACCATGCTCGCTGCGCTCCCAATCTGGCTTGGCTGTGTGCTCTGGA 30781
Oy 3271 CGCTTGAACCTGATGTTTGTAGAAATCATTTTGAACCTTTTCTACAAAGAGGCTT 3330
|||||
Db 30782 CGCTTGAACCTGATGTTTGTAGAAATCATTTTGTAGACCTTTTGTCTACAAAGAGGCTT 30841
Oy 3331 TCTGGAACCTGAGAAAGAAACATCTCTTGCATCTGACACAGTTCTCTACACCATT 3390
|||||

Db 30842 TCTGGACACTGAGAGAACAATCTCTTGGCATTCTGACCAAGTCTCTCTACACATTT 30901
Qy 3391 TTCTCAGCTCATCTTCTGCTGCTGCTCTAAGGAATTTCAATGAGCCTTCTACT 3450
Db 30902 TTCTCAGCTCATCTTCTGCTGCTGCTCTAAGGAATTTCAATGAGCCTTCTACT 30961
Qy 3451 ACTAATTCAGACAGCTTCTCCTCAAAAACCTGTGTGACCTAGCTCTTCAATGACCCTTAACATA 3510
Db 30962 ACTAATTCAGACAGCTTCTCCTCAAAAACCTGTGTGACCTAGCTCTTCAATGACCCTTAACATA 31021
Qy 3511 TGTACATATCTATTAATTTCTGTGTCAATTAATTTTAAAGCAATTAATTC 3570
Db 31022 TGTACATATCTATTAATTTCTGTGTCAATTAATTTTAAAGCAATTAATTC 31081
Qy 3571 CGTTTGGCAAAAGTTAATGATGAGAGCTCTAGAAATCTCAATTTTGGACATATCA 3630
Db 31082 CGTTTGGCAAAAGTTAATGATGAGAGCTCTAGAAATCTCAATTTTGGACATATCA 31141
Qy 3631 GTCTCCTAATATCAGAGATCCCTAAGTCCAGCTGGCTAGTACAGAGTTTTCAGACTT 3690
Db 31142 GTCTCCTAATATCAGAGATCCCTAAGTCCAGCTGGCTAGTACAGAGTTTTCAGACTT 31201
Qy 3691 CCTGCTTCTCAGCTCTCTATTAATCTTAAGACACCAGCATATATCTCTTGAATTAACACC 3750
Db 31202 CCTGCTTCTCAGCTCTCTATTAATCTTAAGACACCAGCATATATCTCTTGAATTAACACC 31261
Qy 3751 TAATGGGAGTGAAGCGAGATGCGACACCTGCGCTGGGGGAGACAGAGTGAAGCT 3810
Db 31262 TAATGGGAGTGAAGCGAGATGCGACACCTGCGCTGGGGGAGACAGAGTGAAGCT 31321
Qy 3811 TTGTCTCTATTAACAAAAGAAAAGAAAAGAAATACAACTTAAGCTCA 3857
Db 31322 TTGTCTCTATTAACAAAAGAAAAGAAAAGAAATACAACTTAAGCTCA 31368

RESULT 4
US-09-880-107-2225
: Sequence 2225, Application US/09880107
: Patent No. US20020142981A1
: GENERAL INFORMATION:
: APPLICANT: Horne, Darci T.
: APPLICANT: Vockley, Joseph G.
: APPLICANT: Scherf, Uwe
: APPLICANT: Gene Logic, Inc.
: TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
: FILE REFERENCE: 44921-5028-WO
: CURRENT APPLICATION NUMBER: US/09/880,107
: CURRENT FILING DATE: 2001-06-14
: PRIOR APPLICATION NUMBER: US 60/211,379
: PRIOR FILING DATE: 2000-06-14
: PRIOR APPLICATION NUMBER: US 60/237,054
: PRIOR FILING DATE: 2000-10-02
: NUMBER OF SEQ ID NOS: 3950
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 2225
: LENGTH: 35641
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: OTHER INFORMATION: Genbank Accession No. US20020142981A1 L76568
US-09-880-107-2225

Query Match 47.4% Score 1829.4: DB 10: Length 35641:
Best Local Similarity 99.4%: Pred. No. 0:
Matches 1836: Conservative 0: Mismatches 11: Indels 0: Gaps 0:

Qy 2011 ACTGACACTGGAAGCCGTGGCCAGAGAACAGATGTGACACAGCAAGCATAGTTGTG 2070
Db 29522 ACTTACTTTTCTCTAGAGTGGCCAGAGAACAGATGTGACACAGCAAGCATAGTTGTG 29581
Qy 2071 GATATGCGTGAATTTGCAAGTGAAGCTTCAATCTCTGATCATGCTGGGGCATTTGACAT 2130
Db 29582 GATATGCGTGAATTTGCAAGTGAAGCTTCAATCTCTGATCATGCTGGGGCATTTGACAT 29641

Qy 2131 GAACCCGAGACTTTAGAGGTTGAGATTAACATCTCAGCAAGAAATGTCGTTGAGGCC 2190
Db 29642 GAACCCGAGACTTTAGAGGTTGAGATTAACATCTCAGCAAGAAATGTCGTTGAGGCC 29701
Qy 2191 AAGAGTATAGATTAATTCGCTCTTAAATTAACGGCGGCTCTACAGCCAGTGATC 2250
Db 29702 AAGAGTATAGATTAATTCGCTCTTAAATTAACGGCGGCTCTACAGCCAGTGATC 29761
Qy 2251 TCCATGTCGGCTACTACAGAGGTCGCCGTCTTGATTTAGTTTGGACCTAGCAAGCCT 2310
Db 29762 TCCATGTCGGCTACTACAGAGGTCGCCGTCTTGATTTAGTTTGGACCTAGCAAGCCT 29821
Qy 2311 TTCTCTCAGCTTCCGAGGTGCTTTTCAAGAGATCTCCAGCAATGACATTAAGTCC 2370
Db 29822 TTCTCTCAGCTTCCGAGGTGCTTTTCAAGAGATCTCCAGCAATGACATTAAGTCC 29881
Qy 2371 AAACCTACTCTTCTTACACTTACTTCCCAAGCATACGGAATCTCTGAGGCCCTCTCC 2430
Db 29882 AAACCTACTCTTCTTACACTTACTTCCCAAGCATACGGAATCTCTGAGGCCCTCTCC 29941
Qy 2431 CATGCAAGCGGAGTTGTTTGAAGAGCTGAACAAGCAAGCACAGCCTGATGGCGG 2490
Db 29942 CATGCAAGCGGAGTTGTTTGAAGAGCTGAACAAGCAAGCACAGCCTGATGGCGG 30001
Qy 2491 ACAGCACTGGCCATTTACAGACAGATTTCTGAACCCCTTCCGAGTCAAGAGATTAATCCT 2550
Db 30002 ACAGCACTGGCCATTTACAGACAGATTTCTGAACCCCTTCCGAGTCAAGAGATTAATCCT 30061
Qy 2551 GGTCCCAAGACTTCTTTTAAAAATGCCAGGGGTGAATGCCAAAAATGCCGCTCTTG 2610
Db 30062 GGTCCCAAGACTTCTTTTAAAAATGCCAGGGGTGAATGCCAAAAATGCCGCTCTTG 30121
Qy 2611 ATGCACACAGTTAAGAATCGCAGAAATTAGCAGCCCTGTCAACAGAGCTACAGAGT 2670
Db 30122 ATGCACACAGTTAAGAATCGCAGAAATTAGCAGCCCTGTCAACAGAGCTACAGAGT 30181
Qy 2671 ATTCTGGGGAATGCTGCAAAATGCCAAACAGCTTATGATTTCAATTCACACCTTTTGA 2730
Db 30182 ATTCTGGGGAATGCTGCAAAATGCCAAACAGCTTATGATTTCAATTCACACCTTTTGA 30241
Qy 2731 GAAGTGTATCAAAAGAAAAGGAAAAGTGAACAGTGAATGCTCTTTCTTATCCAT 2790
Db 30242 GAAGTGTATCAAAAGAAAAGGAAAAGTGAACAGTGAATGCTCTTTCTTATCCAT 30301
Qy 2791 GCCTGTACTTTTACGGGCTCTTGGCAGACATCATAGTCATTAATTAATTTGTTG 2850
Db 30302 GCCTGTACTTTTACGGGCTCTTGGCAGACATCATAGTCATTAATTAATTTGTTG 30361
Qy 2851 CTATTTCAATCTTTTCCAAATGCTCTTAATGATTTGACGGTGAACCCAGGATTC 2910
Db 30362 CTATTTCAATCTTTTCCAAATGCTCTTAATGATTTGACGGTGAACCCAGGATTC 30421
Qy 2911 TCTCTGACCTGACAGTATGAGCATCACTTGAACCTGCTGCTCTCTTTTCCCTCC 2970
Db 30422 TCTCTGACCTGACAGTATGAGCATCACTTGAACCTGCTGCTCTCTTTTCCCTCC 30481
Qy 2971 TGCACGCTATGCGGGCTTACAGATGTTCTTTTAAATGAGGTGTGTGAGATCAGT 3030
Db 30482 TGCACGCTATGCGGGCTTACAGATGTTCTTTTAAATGAGGTGTGTGAGATCAGT 30541
Qy 3031 AAAGTTCTACAAAGTATACAGAGGTAGCAAACTTACCTGATCTTAACAGATCTAT 3090
Db 30542 AAAGTTCTACAAAGTATACAGAGGTAGCAAACTTACCTGATCTTAACAGATCTAT 30601
Qy 3091 TAGAAAGGAATATGCTTAAGCCTGGCATGAGCGGTGCAAGGGGAGGAAAAGAGAGCA 3150
Db 30602 TAGAAAGGAATATGCTTAAGCCTGGCATGAGCGGTGCAAGGGGAGGAAAAGAGAGCA 30661
Qy 3151 GAAAGTACCATTTTAAAGAGTCTTGTATCTAGTCAACATTAATTAAGAGTCTTAAT 3210
Db 30662 GAAAGTACCATTTTAAAGAGTCTTGTATCTAGTCAACATTAATTAAGAGTCTTAAT 30721

OY	3211	GCCTTATACCATTGCTGGGCTCCCAATCGAGTCCTTGCTGGTGTGCTGGTA	3270
Db	30722	GCACCTATAACCATGTCCTCTGGCTCTCCAATTCGGTCTTGCTGTGTCTGGTA	30781
OY	3271	GCCTTGAACGTATGTTGTGTAGAATAATCATGTTCTGACCCCTTGTCTACAAGAGCCT	3330
Db	30782	CGCTTAGAATCATGTTGTGTAGAATAATCATGTTCTGACCCCTTGTCTACAAGAGCCT	30841
OY	3331	TCTGGAACTGAGAGAAACAATCTCTTGGCATTCCTGACAGAGTTCTCTACACAT	3390
Db	30842	TCTGGAACTGAGAGAAACAATCTCTTGGCATTCCTGACAGAGTTCTCTACACAT	30901
OY	3391	TTCTTCACGCTCATCTACTCTGCTGCTCTGCTCTAGAGAAATTCATGAGCCTTCTACT	3450
Db	30902	TTCTTCACGCTCATCTACTCTGCTGCTCTGCTCTAGAGAAATTCATGAGCCTTCTACT	30961
OY	3451	ACTAATTCAGACAGCTCTCTCAAAAACCTGGTGTGACTAGTCTCTTAATGACCTTAACATA	3510
Db	30962	ACTAATTCAGACAGCTCTCTCAAAAACCTGGTGTGACTAGTCTCTTAATGACCTTAACATA	31021
OY	3511	TGTACATATATCTAATTAATTCATGTTGCCAATTAATTAATTTTAAAGCAAATGAAATTC	3570
Db	31022	TGTACATATATCTAATTAATTCATGTTGCCAATTAATTAATTTTAAAGCAAATGAAATTC	31081
OY	3571	CTGTTTGCAAAAGTTAATGATGAGAGAGCTTTAGAATTCATATTTTGCACATATTTCA	3630
Db	31082	CTGTTTGCAAAAGTTAATGATGAGAGAGCTTTAGAATTCATATTTTGCACATATTTCA	31141
OY	3631	GTCCTCTAATATCAAGATCCCTAGTCCAGCTGGCTAGTTTACAGATTTTTCAGACT	3690
Db	31142	GTCCTCTAATATCAAGATCCCTAGTCCAGCTGGCTAGTTTACAGATTTTTCAGACT	31201
OY	3691	CCTGCTTCTCAGCTCTTATATCTTAATCTTAAGCACACACATATATCTCTGAATAACAAC	3750
Db	31202	CCTGCTTCTCAGCTCTTATATCTTAATCTTAAGCACACACATATATCTCTGAATAACAAC	31261
OY	3751	TAAATGGCAGTAGCCGAGATTCGACACACCTGACCCCTGCTGGCGACAGAGTAGACT	3810
Db	31262	TAAATGGCAGTAGCCGAGATTCGACACACCTGACCCCTGCTGGCGACAGAGTAGACT	31321
OY	3811	TTGCTCTAATTCAAAAAGAAAAGAAATTAACAACTAAGCTCA 3857	
Db	31322	TTGCTCTAATTCAAAAAGAAAAGAAATTAACAACTAAGCTCA 31368	
 RESULT 5 US-09-867-701-10472/c			
; Sequence 10472, Application US/09867701			
; Patent No. US20020132237A1			
; GENERAL INFORMATION:			
; APPLICANT: Agilate, Paul A.			
; APPLICANT: Jones, Robert			
; APPLICANT: Harlocker, Susan L.			
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY			
; TITLE OF INVENTION: AND DIAGNOSIS OF OVARIAN CANCER			
; FILE REFERENCE: 210121.497			
; CURRENT APPLICATION NUMBER: US/09/867,701			
; CURRENT FILING DATE: 2001-05-29			
; NUMBER OF SEQ ID NOS: 10912			
; SOFTWARE: FastSeq for Windows Version 4.0			
; SEQ ID NO 10472			
; LENGTH: 601			
; TYPE: DNA			
; ORGANISM: Homo sapien			
; FEATURE:			
; NAME/KEY: misc_feature			
; LOCATION: (1)...(601)			
; OTHER INFORMATION: n = A,T,C or G			
US-09-867-701-10472			

Query Match	6.6%;	Score 256.2;	DB 10;	Length 601;
Best Local Similarity	97.7%;	Pred. No. 1.2e-53;		
Matches 258;	Conservative 0;	Mismatches 6;	Indels 0;	Gaps 0;

Qy	1561	GATTATGACAGTATGCCAGAAACCTGCCCCGAGAAATTAACATCAAGAAATTGATGTA	1620
Db	601	GAAATACCCAGTATGCCAGAAACCTGCCCCGAGAAATTAACATCAAGAAATTGATGTA	542
Qy	1621	AATTGTGATCGATGCTGCTTTCCGGAATCTGAAAGAACCCCTACTATCATCTCCATCG	1680
Db	541	AATTGTGATCGATGCTGCTTTCCGGAATCTGAAAGAACCCCTACTATCATCTCCATCG	482
Qy	1681	CTTCTGGGTTGCAGCCAGACCCCTATGCTCTGCACAAAGGTACTACATCAATGATGAGCCAGA	1740
Db	481	CTTCTGGGTTGCAGCCAGACCCCTATGCTCTGCACAAAGGTACTACATCAATGATGAGCCAGA	422
Qy	1741	TACGTGGTCTTTTATGACGAGAGACTAACTTTGTTCGGCAGCTTGAATTTACAGGGCG	1800
Db	421	TACGTGGTCTTTTATGACGAGAGACTAACTTTGTTCGGCAGCTTGAATTTACAGGGCG	362
Qy	1801	ACTAGGCTGGGAACCTCTGAGG	1824
Db	361	ACTAGGCTGGGAACCTCTGAGG	338

```

: RESULT 6
: US-09-919-497-14/C
: Sequence 14, Application US/09919497
: Patent No. US2002010662A1
: GENERAL INFORMATION:
: APPLICANT: Mutter, George L.
: TITLE OF INVENTION: PROGNOSTIC CLASSIFICATION OF ENDOMETRIAL CANCER
: FILE REFERENCE: B0801/7225
: CURRENT APPLICATION NUMBER: US/09/919,497
: CURRENT FILING DATE: 2001-07-31
: PRIOR APPLICATION NUMBER: US 60/221,735
: PRIOR FILING DATE: 2000-07-31
: NUMBER OF SEQ ID NOS: 100
: SOFTWARE: PatentIn version 3.0
: SEQ ID NO 14
: LENGTH: 3857
: TYPE: DNA
: ORGANISM: Homo sapiens
: US-09-919-497-14

```

Query Match	4.08;	Score 154.6;	DB 10;	Length 3857;
Best Local Similarity	82.08;	Pred. No. 3.9e-28;		
Matches 178; Conservative	0;	Mismatches 39;	Indels 0;	Gaps 0;

QY	2941	AAATGGCGTGGCCGCGCTCTTTTTCGCGCCGAGCGCTATGACCGGGCTTAGCAATGTT	3000
Db	3157	AGCTTCTTGCGCTGCTCTTTTCCCTCCCTGACCGCTCATGCGAGCTTAGCAATTC	30989
QY	3001	CTTTTAAATAGCTTTTGTCAGCATCAGTAAGTTCTCTACAAAGTATATACAAAGTAG	3060
Db	3097	CTTCTCAATATAGATTCGTGTAGATCAAGTAAGTTCTTACCTCTGTATATCACTGTAG	3038
QY	3061	AAACTTACCTGATCCTPACAGATCTCATTTAGAAAGAAATATGCTAACCTGGCATGGA	3120
Db	3037	GAACCTTACCTGATCTCGACAAACCTCATTTAAAAAGAAACATGCTAACCCGCGCATAGA	2978
QY	3121	CGGTCCAGGGAGGAAAAAGACAGCGCAAGAAGCT	3157
Db	2977	CGGTCCAGGGAGGAAAAAGACAGCGCAAGAAGTT	2941

```

; RESULT 7
; US-09-962-436-306/C
; Sequence 306, Application US/09962436
; Patent No. US20020081301A1
;
; GENERAL INFORMATION:
; APPLICANT: Soppet, Daniel
; TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using Sigm
; TITLE OF INVENTION: Sets
; FILE REFERENCE: 689290-75
; CURRENT APPLICATION NUMBER: US/09/962,436

```

```

: CURRENT FILING DATE: 2001-09-25
: PRIOR APPLICATION NUMBER: US/60/235,082
: PRIOR FILING DATE: 2000-09-25
: PRIOR APPLICATION NUMBER: US/60/234,924
: PRIOR FILING DATE: 2000-09-25
: NUMBER OF SEQ ID NOS: 568
: SOFTWARE: PatentIn version 3.0
: SEQ ID NO 306
: LENGTH: 35641
: TYPE: DNA
: ORGANISM: Homo sapiens
: US-09-962-436-306

```

Query Match	4.08;	Score 154.6;	DB 10;	Length 35641;
Best Local Similarity	82.08;	Pred. NO. 1.3e-27;		
Matches 178; Conservative	0;	Mismatches 39;	Indels 0;	Gaps 0;

QY	2941	AAATGGCCGTCGCCGTCCTTTTCTCCGCTCCACCGTCCTAAGCCGGGCTTAGCAATGTT	3000
Db	30668	AGCTTCTCTTGTCCTGCTCTTTTCCCTCCGCTACCGCTCCAGGCTTAGCAATTC	30609
QY	3001	CTTTTAAATGAGCTTTGTGCAGGATCAGGTAAGTCTCTACAAGATTTACAGAAGTGG	3060
Db	30608	CTTCTTAATAGATCTGTGTAGATCAGGTAAAGTTTCTACTTCTGTATACTTTG	30549
QY	3061	AAACTTTACCTGATCCTTAACAGATCTCATTTTGAAGAAGATATGCTTAAGCCCTGGCATGA	3120
Db	30548	GAACCTTTACCTGATCCTGTGACAACTCATTTTAAAAAAGAACATGCTAAGCCCGCATGA	30489
QY	3121	CGGTGACAGGAGGAAAAAGAGCAGCCACAAAGAACT	3157
Db	30488	CGGTGACAGGAGGAAAAAGAGCAGCCACAAAGAACT	30452

RESULT 8
US-09-880-107-2225/c

sequence: 2222, Application US/039601/
 Patent No. US20020142981A1
 GENERAL INFORMATION:
 APPLICANT: Horne, Darci T.
 APPLICANT: Voelckley, Joseph G.
 APPLICANT: Scherf, Uwe
 APPLICANT: Gene Logic, Inc.
 TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
 FILE REFERENCE: 44921-5028-WO
 CURRENT APPLICATION NUMBER: US/09/880,107
 CURRENT FILING DATE: 2001-06-14
 PRIOR APPLICATION NUMBER: US 60/211,379
 PRIOR FILING DATE: 2000-06-14
 PRIOR APPLICATION NUMBER: US 60/237,054
 PRIOR FILING DATE: 2000-10-02
 NUMBER OF SEQ ID NOS: 3950
 SOFTWARE: PatentIn Ver. 2.1
 SEQ ID NO 2225
 LENGTH: 35641
 TYPE: DNA
 ORGANISM: Homo sapiens
 FEATURE:
 OTHER INFORMATION: Genbank Accession No. US20020142981A1 L76568
 US-09-880-107-2225

Query Match	4.0%;	Score 154.6;	DB 10;	Length 35641;
Best Local Similarity	82.0%;	Pred. No. 1.3e-27;		
Matches 178;	Conservative	0;	Mismatches 39;	Indels 0;
				Gaps 0;

QY	2941	AATTTGCGTGGCCGTCCTTTTCTCCTCCGCACAGTGCTATGCCGGGTGTAAGCATGTTT	3000
Dd	30668	AGCTTTCTCTTGCCCTGCTCTTTTTCCCCTCCACACGGTCACATGCCCCAGGCTTAGCATATTCC	30690
QY	3001	CTTTTAAATGAGGTTTCTCAGGATCAGSTAAGTTCTCTACCAAGTATTACAGAAGTAG	3060
bD	30608	CTTTCTAATATGAGACTCTTTAGGATCAGATAAAGTTTCTTACCCTCTTGTAATCACTTTGAG	30549

QY	3061	AACTTTACCGATTCCTAACGATTCCTCAATTTGAAGAATATGC	TAAGCCCGCATGGA	3120
Db	30548	GAACTTTACCTGATCCTCGCAAAACCTCATTTTAAAGAAGACATGCT	TAAGCCCGCATTGA	304899
QY	3121	CGGTCCAGGAGGAAAAAGCAGCGACACAAAGCT		3157
Db	30488	CGGTCCAGGAGGAAAAAGCAGCGACACAGCAATTT		30452

RESULT 9
US-09-764-864-1768/c
; Sequence 1768, Application US/09764864

```

; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic acids, proteins, and Antibodies
; FILE REFERENCE: P1723
; CURRENT APPLICATION NUMBER: US/09/764,864
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1792
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1768
; LENGTH: 9566
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-764-864-1768

```

Query Match	1.88;	Score 70.2;	DB 10;	Length 9566;
Best Local Similarity	85.78;	Pred. No. 4.3e-07;		
Matches 78;	Conservative 0;	Mismatches 13;	Indels 0;	Gaps 0;

[illegible]

RESULT 10
US-09-764

```

? sequence 4488, application US/09/04867/
? Patent No. US20020147140A1
? GENERAL INFORMATION:
? APPLICANT: Rosen et al.
? TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
? FILE REFERENCE: PC005
? CURRENT APPLICATION NUMBER: US/09/764,877
? CURRENT FILING DATE: 2001-01-17
? Prior application data removed - refer to PALM or file wrapper
? NUMBER OF SEQ ID NOS: 4031
? SOFTWARE: patentin Ver. 2.0
? SEQ ID NO 3486
? LENGTH: 9566
? TYPE: DNA
? ORGANISM: Homo sapiens
? US-09-764-877-3486

```

Query Match	1.8%;	Score 70.2;	DB 10;	Length 9566;
Best Local Similarity	85.7%;	Pred. No. 4.3e-07;		
Matches 78;	Conservative 0;	Mismatches 13;	Indels 0;	Gaps 0;

[illegible]

```

RESULT 11
US-09-764-877-2510
; Sequence 2510, Application US/09764877
; Patent No. US20020147140A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC005
; CURRENT APPLICATION NUMBER: US/09/764,877
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 4031
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2510
; LENGTH: 3705
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-877-2510

Query Match 1.8%; Score 69.2; DB 10; Length 3705;
Best Local Similarity 85.6%; Pred. No. 4.5e-07;
Matches 77; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

OY 3755 TGGCAGTAGGCCGAGATCGACCACTGCACCCCTGCGGGCGACAGAGTAGACTTGT 3814
Db 1900 TTGCAGTAGGCCAGAGATCGACCACTGCACCTCGGCGACACAGTAGACTCTGT 1959

OY 3815 CTCTATTACAAAAGAAAGAAAGAAAGAAATA 3844
Db 1960 CTCAAAAAAGAAAGAAAGAAAGAAAGATGTA 1989

RESULT 12
US-09-969-708-79
; Sequence 79, Application US/09969708
; Patent No. US20020102532A1
; GENERAL INFORMATION:
; APPLICANT: Augustus, Meena
; TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using Signature
; FILE REFERENCE: 689290-70
; CURRENT APPLICATION NUMBER: US/09/969,708
; CURRENT FILING DATE: 2001-10-03
; PRIOR APPLICATION NUMBER: US/60/237,606
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: US/60/237,608
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: US/60/237,425
; PRIOR FILING DATE: 2000-10-03
; NUMBER OF SEQ ID NOS: 658
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 79
; LENGTH: 145831
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-969-708-79

Query Match 1.8%; Score 69.2; DB 10; Length 145831;
Best Local Similarity 81.6%; Pred. No. 3.4e-06;
Matches 80; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

OY 3755 TGGCAGTAGGCCGAGATCGACCACTGCACCCCTGCGGGCGACAGAGTAGACTTGT 3814
Db 38704 TTGCAGTAGGCCAGAGATCGACCACTGCACCTCGGCGACAGAGTAGACTCTGT 38763

OY 3815 CTCTATTACAAAAGAAAGAAAGAAAGAAATACACCTAA 3852
Db 38764 CTCAAAAAAGAAAGAAAGAAAGAAAGAAAGTGA 38801

RESULT 13
US-09-954-456-2116
; Sequence 2116, Application US/09954456

```

```

Patent No. 020020115057A1
GENERAL INFORMATION:
APPLICANT: Young, Paul
TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using C
FILE REFERENCE: 689290-76
CURRENT APPLICATION NUMBER: US/09/954,456
CURRENT FILING DATE: 2001-09-18
PRIORITY APPLICATION NUMBER: US/60/233,617
PRIORITY FILING DATE: 2000-09-18
PRIORITY APPLICATION NUMBER: US/60/234,052
PRIORITY FILING DATE: 2000-09-20
PRIORITY APPLICATION NUMBER: US/60/234,923
PRIORITY FILING DATE: 2000-09-25
PRIORITY APPLICATION NUMBER: US/60/235,134
PRIORITY FILING DATE: 2000-09-25
PRIORITY APPLICATION NUMBER: US/60/235,637
PRIORITY FILING DATE: 2000-09-26
PRIORITY APPLICATION NUMBER: US/60/235,638
PRIORITY FILING DATE: 2000-09-26
PRIORITY APPLICATION NUMBER: US/60/235,711
PRIORITY FILING DATE: 2000-09-27
PRIORITY APPLICATION NUMBER: US/60/235,720
PRIORITY FILING DATE: 2000-09-27
PRIORITY APPLICATION NUMBER: US/60/235,863
NUMBER OF SEQ ID NOS: 2276
SOFTWARE: PatentIn version 3.0
SEQ ID NO 2116
LENGTH: 145831
TYPE: DNA
ORGANISM: Homo sapiens
US-09-954-456-2116

Query Match      1.8%; Score 69.2; DB 10; Length 145831;
Best Local Similarity 81.6%; Pred. No. 3.4e-06;
Matches 80; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

QY   3755 TGGCGATGAGCCGAGATGCACACCCTGCCCTGGCGCAGCATGTGACTTTGT 3814
      ||||| ||| ||||||||| ||||||| ||| ||||| ||||| ||||| |||
DB  38704 TTGCATTAGAGAGATGCACCTGCATCTCCAGCTTGGTGACACGCAAGACTCTGT 38763
QY   3815 CTCTATTACAAAGAAGAAAAGAAATAATACACCTAA 3852
      ||| | ||||| ||||||||| ||| ||
Db   38764 CTCAAAAAAGAAAAAGAAAAGAAAAGAAACTGAA 38801

RESULT 14
US-09-860-670-263
: Sequence 263, Application us/09860670
: Patent No. US20020165137A1
GENERAL INFORMATION:
APPLICANT: Ruben et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: PA127P1
CURRENT APPLICATION NUMBER: US/09/860,670
CURRENT FILING DATE: 2001-05-21
Prior application data removed - consult PALM or file wrapper
NUMBER OF SEQ ID NOS: 289
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 263
LENGTH: 24218
TYPE: DNA
ORGANISM: Homo sapiens
US-09-860-670-263

Query Match      1.8%; Score 69; DB 9; Length 24218;
Best Local Similarity 83.9%; Pred. No. 1.4e-06;
Matches 78; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY   3755 TGGCAGTAGCCGAGATGCACACCCTGCCCTGGCGCAGCATGTGAGACTTTGT 3814

```

Db 6215 TTGCAGTGAGCCGAGATCGCACCCTGCGACTCTCCGCTGGCGACAGATGAGACTCCGT 6274

Oy 3815 CTCTATTACAAAAGAAAAGAAAAGAAATACA 3847

Db 6275 CTCAAAACAAAAAAGAAAAAAGAAAAA 6307

RESULT 15

US-09-764-410

; Sequence 410, Application US/09764878

; Patent No. US2002090615A1

; GENERAL INFORMATION:

; APPLICANT: Rosen et al.

; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies

; FILE REFERENCE: PA121

; CURRENT APPLICATION NUMBER: US/09/764,878

; PRIOR APPLICATION DATE: 2001-01-17

; PRIOR APPLICATION data removed - consult PALM or file wrapper

; NUMBER OF SEQ ID NOS: 428

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 410

; LENGTH: 323

; TYPE: DNA

; ORGANISM: Homo sapiens

US-09-764-878-410

Query Match

Best Local Similarity 1.8%; Score 68; DB 10; Length 323;

Matches 77; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

Oy 3755 TGGCAGTGAGCCGAGATCGCACCCTGCGCTGGCGACAGATGAGACTTTGT 3814

Db 219 TTGCAGTGAGCCGAGATCGCACCCTGCGACTCTCCGCTGGCGACAGATGAGACTCTGT 278

Oy 3815 CTCTATTACAAAAGAAAAGAAAAGAAATACA 3846

Db 279 CTCAAAACAAAAAAGAAAAAAGAGA 310

Search completed: November 26, 2002, 23:47:25

Job time : 1153.71 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 26, 2002, 19:44:35 : Search time 1162.24 Seconds

(without alignments)
16827.091 Million cell updates/sec

Title: US-09-919-497-32

Perfect score: 672

Sequence: 1 aataaactttctttt.....tttttttttttgaga 672

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: gb_da:*
2: gb_hlg:*
3: gb_in:*
4: gb_cm:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_da:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_hlg_hum:*
31: em_hlg_inv:*
32: em_hlg_other:*
33: em_hlg_mus:*
34: em_hlg_pln:*
35: em_hlg_rnd:*
36: em_hlg_mam:*
37: em_hlg_vrt:*
38: em_sy:*
39: em_hlgo_hum:*
40: em_hlgo_mus:*
41: em_hlgo_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	672	100.0	672	9 HUMGA7A08	M93036 Human (clon
2	672	100.0	171987	9 AC0792775	AC0792775 Homo sapi
3	461.6	68.7	621	6 AX197746	AX197746 Sequence
4	461.6	68.7	621	6 AX208351	AX208351 Sequence
5	451.4	67.2	1545	6 AX014895	AX014895 Sequence
6	449.8	66.9	1565	6 AK026585	AK026585 Homo sapi
7	449.4	66.9	1504	9 HSKSA	X14758 Human mRNA
8	449.4	66.9	1504	9 HUMKSA	M32325 Human adeno
9	447.8	66.6	1457	9 HUMGA7A	M33011 Human (clon
10	443.4	66.0	637	6 AX395799	AX395799 Sequence
11	442.4	65.8	481	6 AX397344	AX397344 Sequence
12	442.2	65.8	545	6 AX198887	AX198887 Sequence
13	439.4	65.4	1511	9 HUMMABK514	M26481 Homo sapien
14	439.4	65.3	439	6 AX209414	AX209414 Sequence
15	438.4	65.2	1372	9 HUMCSAE	M32306 Human epith
16	415.4	61.8	425	6 AX193024	AX193024 Sequence
17	362.4	53.9	1458	9 BC014785	BC014785 Homo sapi
18	342.6	51.0	173814	9 AC093840	AC093840 Homo sapi
19	335.8	50.0	507	6 AX209701	AX209701 Sequence
20	333.8	49.7	555	6 AX395792	AX395792 Sequence
21	326	48.5	326	6 AX396230	AX396230 Sequence
22	269	40.0	269	6 AX192570	AX192570 Sequence
23	234	34.8	235	11 G43776	G43776 WIAF-3376-S
24	222.8	33.2	272	6 AX182073	AX182073 Sequence
25	222	33.0	353	6 AX351467	AX351467 Sequence
26	221.2	32.9	223	6 AX340829	AX340829 Sequence
27	221.2	32.9	223	6 AX341705	AX341705 Sequence
28	220.2	32.8	223	6 AX340713	AX340713 Sequence
29	214	31.8	214	6 I06778	I06778 Sequence 4
30	199.2	29.6	201	6 AX351491	AX351491 Sequence
31	185.4	27.6	691	6 AX351469	AX351469 Sequence
32	183.6	27.3	238	6 AX340174	AX340174 Sequence
33	183.4	27.3	554	6 AX192846	AX192846 Sequence
34	182.4	27.1	205	6 AX341377	AX341377 Sequence
35	178	26.5	178	6 AX397271	AX397271 Sequence
36	171.6	25.5	462	6 AX339773	AX339773 Sequence
37	145	21.6	160	6 AX197888	AX197888 Sequence
38	145	21.6	160	6 AX208485	AX208485 Sequence
39	112	16.7	191914	2 AC104671	AC104671 Mus muscu
40	107	15.9	174433	2 AC101909	AC101909 Mus muscu
41	96.6	14.4	84604	10 AL645763	AL645763 Mouse DNA
42	93.8	14.0	1492	10 BC005618	BC005618 Mus muscu
43	91.2	13.6	1710	6 AX305882	AX305882 Sequence
44	91.2	13.6	1710	10 M0SEBP	M76124 Mus musculu
45	86.2	12.8	1524	10 RNEGP314H	AJ001044 Rattus no

ALIGNMENTS

RESULT 1
LOCUS HUMGA7A08
DEFINITION Human (clone 21726) carcinoma-associated antigen GA733-2 (GA733-2)
ACCESSION M93036
VERSION M93036.1 GI:182904
KEYWORDS carcinoma-associated antigen GA733-2.
SEGMENT 8 of 8
SOURCE Homo sapiens (tissue library: UN67) Lymphoma DNA.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (sites)
AUTHORS Szala,S., Froehlich,M., Scollon,M., Kasai,T., Steplewski,Z.,

Koprowski, H. and Linnenbach, A.J.
Molecular cloning of cDNA for the carcinoma-associated antigen GA733-2
Proc. Natl. Acad. Sci. U.S.A. 8, 3542-3546 (1990)
2 (bases 1 to 672)

Linnebach, A.J., Seng, B.A., Wu, S., Robbins, S., Scollon, M., Pyrc, J.J., Druck, T. and Huebner, K.
Retroposition in a family of carcinoma-associated antigen genes
Mol. Cell. Biol. 13 (3), 1507-1515 (1993)
JOURNAL 93180797
MEDLINE 93180797
PUBMED 8382772

FEATURES
source Location/Qualifiers
1..672
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="iv"
/cell_type="B-cell"
/tissue_type="lymphoma"
/tissue_id="LN87"
join(M93029,1:35..573,M93030,1:1..759,M93031,1:1..355,M93032,1:1..367,M93033,1:1..596,M93034,1:1..568,M93035,1:1..336,1..606)
/gene="GA733-2"
join(M93029,1:269..522,M93030,1:83..190,M93030,1:428,M93031,1:30..95,M93032,1:163..226,M93033,1:304..405,M93034,1:265..465,M93035,1:95..139,160..606)
/gene="GA733-2"
join(M93029,1:447..532,M93030,1:83..190,M93030,1:428,M93031,1:30..95,M93032,1:163..226,M93033,1:304..405,M93034,1:265..465,M93035,1:95..139,160..201)
/gene="GA733-2"
/codon_start=1
/product="carcinoma-associated antigen GA733-2"
/protein_id="AAB0075.1"
/db_xref="GI:182906"
/translation="MAPPOVLAFLGLLAATRTFAAOEBCVENCYKALVNCFVNMR
OCQCTSVGAQNTVICSKLAAKCLVMAKENGSKLRAPREGALDNNGLDPCDES
GLFKAKOCNGSTCVNTAGVTRDKDETSERVRYYWIIELKHAREKPYDSK
SLRTALOKETITRQLDPEFTISILYENNVTIDLVONSCKTDNDVADVAAYFEK
DYKGSLEFHSKMKMLVNGEOLDLPGOTLIYVDEKAEFEFMGLKAGVIAVIVVV
IAVAGIVIVLVSRKKRAKYEKAEIKEMGHRRLNM"
order(M93035,1:140..336,1..159)
/gene="GA733-2"
/number=8
160..606
/gene="GA733-2"
/number=9
500..505
/gene="GA733-2"
518
/gene="GA733-2"
588..593
/gene="GA733-2"
606
polyA_site /gene="GA733-2"
BASE COUNT 222 a 85 c 100 g 265 t
ORIGIN

Query Match 100.0%; Score 672; DB 9; Length 672;
Best Local Similarity 100.0%; Pred. No. 2.7e-101;
Matches 672; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AATATACTTTTCTTTTATTAATAAGCTTAATAATGGGAAAAATATCTT 60
DB 1 AATATACTTTTCTTTTATTAATAAGCTTAATAATGGGAAAAATATCTT 60
QY 61 GGTTCCTTTAATTCATTTTATTAATACATATTTTCAGAAATGAAAGATTGAAA 120
DB 61 GGTTCCTTTAATTCATTTTATTAATACATATTTTCAGAAATGAAAGATTGAAA 120
QY 121 ATATATTAGAAATTTTTCGTGCTTTTCCCTTTTCAGATAAGAGATGGTGAGTG 180
DB 121 ATATATTAGAAATTTTTCGTGCTTTTCCCTTTTCAGATAAGAGATGGTGAGTG 180

QY 181 CATAGGAACCTCATGCTATATATATTTGAAGATTATAGAA..AAGGAATAGCAA 240
DB 181 CATAGGAACCTCATGCTATATATATTTGAAGATTATAGAA..AAGGAATAGCAA 240
QY 241 ATGACACAAATATACAAATGTGTGTCGCGAGCAAGACATCTTGAAGCTCATAGTT 300
DB 241 ATGACACAAATATACAAATGTGTGTCGCGAGCAAGACATCTTGAAGCTCATAGTT 300
QY 301 TGTATTACATCATATATTTTGTAAATAGTGAACCTGTACTCAAAATATAGACGTT 360
DB 301 TGTATTACATCATATATTTTGTAAATAGTGAACCTGTACTCAAAATATAGACGTT 360
QY 361 GAAACGTGCTTACCAATCTGAAATTTGACCAACGAGTCTATATATGAGATATAT 420
DB 361 GAAACGTGCTTACCAATCTGAAATTTGACCAACGAGTCTATATATGAGATATAT 420
QY 421 GTAAATCCAGAACTTGACCTCATCGTTAAATTTATATGTATGTAACATTCATGTGT 480
DB 421 GTAAATCCAGAACTTGACCTCATCGTTAAATTTATATGTATGTAACATTCATGTGT 480
QY 481 GCATTAAATATGCTTCCACAGTAATAATTCGAAAACTGATTTGTGATGAAAGCTGCTT 540
DB 481 GCATTAAATATGCTTCCACAGTAATAATTCGAAAACTGATTTGTGATGAAAGCTGCTT 540
QY 541 TCTATTACTTGAGTCTTGATACATACATCTTTTATAGAGTATGAAATTAACAATTTT 600
DB 541 TCTATTACTTGAGTCTTGATACATACATCTTTTATAGAGTATGAAATTAACAATTTT 600
QY 601 AAACGTAAATTTCTTAACCTTGACATTTCAAAATTCCTCTCTCTCTCTCTCTCTCT 660
DB 601 AAACGTAAATTTCTTAACCTTGACATTTCAAAATTCCTCTCTCTCTCTCTCTCTCT 660
QY 661 TTTTCTTTGAGA 672
DB 661 TTTTCTTTGAGA 672

RESULT 2
AC079775 171987 bp DNA linear PRI 29-MAY-2002
LOCUS Homo sapiens BAC clone RP11-295P2 from 2, complete sequence.
DEFINITION AC079775
AC079775
AC079775.6 GI:19848453
VERSION
KEYWORDS
SOURCE
ORGANISM
Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE
AUTHORS Sulston, J.E. and Waterston, R.
TITLE Toward a complete human genome sequence
JOURNAL Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE 98063792
PUBMED 9847074
REFERENCE
AUTHORS Belter, E., Haekenson, W., Doeber, A. and Elliott, G.
TITLE The sequence of Homo sapiens BAC clone RP11-295P2
JOURNAL Unpublished (2001)
REFERENCE
AUTHORS Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (10-SEP-2000) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
4 (bases 1 to 171987)
REFERENCE
AUTHORS Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (30-MAR-2002) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
5 (bases 1 to 171987)
REFERENCE
AUTHORS Waterston, R.

TITLE Direct Submission
JOURNAL Submitted (29-MAY-2002) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT On Mar 30, 2002 this sequence version replaced gi:16924154.

----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: sapiens@wustl.wustl.edu
----- Summary Statistics
Center project name: H_NH0295P02

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions:

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:
Mapping information for this clone was provided by Dr. John D. McHersom, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see http://genome.wustl.edu/gsc

SOURCE INFORMATION:
The RP11 human BAC library was made from the blood of one male donor, as described by Osoegawa, K., Woon, P. Y., Zhao, B., Frengen, E., Tateo, M., Catanesi, J. J. and de Jong, P. J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong and coworkers at http://www.chori.org

NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the left is RP11-261E8; the clone sequenced to the right is RP11-436K12. Actual start of this clone is at base position 1 of RP11-295P2; actual end is at base position 171987 of RP11-295P2.

FEATURES

source

1. 171987
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="2"
/map="2"
/clone="RP11-295P2"
/clone_id="RP11-11"
496. 539
/rpt_family="MERL_type"
540. 840
/rpt_family="Alu"
841. 1002
/rpt_family="MERL_type"
1003. 1306
/rpt_family="Alu"
1307. 1369
/rpt_family="MERL_type"
1703. 1951
/note="match to EST BE504918 (NID:g9707326) hz32e08.x1"
1770. 2076
/rpt_family="Alu"
2455. 2832
/note="similar to Homo sapiens EST BF819288 (NID:g12157087)"
repeat_region 2612. 2687

/rpt_family="MIR"
2688. 2989
/rpt_family="Alu"
2990. 3052
/rpt_family="MIR"
3268. 3287
/rpt_family="(A)n"
3305. 3412
/rpt_family="L1"
3577. 3886
/rpt_family="Alu"
4004. 4321
/rpt_family="Alu"
4322. 4626
/rpt_family="ERV1"
4629. 4703
/rpt_family="Alu"
4669. 4695
/note="match to EST BG982215 (NID:g14384950)"
5895. 5994
/rpt_family="MIR"
6612. 6975
/rpt_family="MALR"
7608. 8109
/note="match to EST BF105129 (NID:g10887655)"
7625. 7727
/rpt_family="MIR"
7735. 7876
/rpt_family="MERL_type"
8217. 8547
/rpt_family="L2"
8984. 9245
/rpt_family="Alu"
9251. 9279
/rpt_family="(CAA)n"
9283. 9334
/rpt_family="Alu"
9335. 9640
/rpt_family="Alu"
9641. 9800
/rpt_family="Alu"
9802. 10082
/rpt_family="Alu"
10088. 10118
/rpt_family="(TAA)n"
10385. 10465
/rpt_family="MIR"
10667. 10961
/rpt_family="Alu"
11074. 11530
/rpt_family="MALR"
11730. 12034
/rpt_family="Alu"
12039. 12322
/rpt_family="Alu"
12372. 12408
/rpt_family="(TCCC)n"
12461. 12651
/rpt_family="MIR"
12747. 13036
/rpt_family="Alu"
13097. 13413
/rpt_family="Alu"
13445. 13730
/rpt_family="Alu"
13841. 13968
/rpt_family="MIR"
13969. 14260
/rpt_family="Alu"
14261. 14368
/rpt_family="MIR"
14412. 14720
/rpt_family="Alu"
repeat_region

repeat_region 15009..15312
/rpt_family="Alu"
misc_feature 15796..16451
/note="match to EST AM853926 (NID:g7949619)"
repeat_region 16612..16728
/rpt_family="MIR"
repeat_region 17083..17384
/rpt_family="Alu"
misc_feature 17168..17169
/note="match to EST AM853926 (NID:g7949619)"
repeat_region 17563..17713
/rpt_family="MIR"
repeat_region 17716..18009
/rpt_family="Alu"
repeat_region 18010..18118
/rpt_family="(CACTA)n"
misc_feature 18199..18502
/note="match to EST BE504918 (NID:g9707326) h32608.x1"
repeat_region 18641..18855
/rpt_family="L2"
repeat_region 19090..19399

Query Match 100.0%; Score 672; DB 9; Length 171987;
Best Local Similarity 100.0%; Pred. No. 7.5e-102;
Matches 672; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AATATACCTTTCTTTTATTAATTAAGCTTTAATGTTGAGGAAAAATTTACTT 60
DB 81402 AATATACCTTTCTTTTATTAATTAAGCTTTAATGTTGAGGAAAAATTTACTT 81461

QY 61 GTGTCTTTAAATTCATTTTATTAATCTATTTTCAGATGACAAAGATTGAAA 120
DB 814b2 GTGTCTTTAAATTCATTTTATTAATCTATTTTCAGATGACAAAGATTGAAA 81521

QY 121 ATTATTAGAATTTTCTGTGCTTTTCCGTTCAGATTAAGSAGATGGGATG 180
DB 81522 ATTATTAGAATTTTCTGTGCTTTTCCGTTCAGATTAAGSAGATGGGATG 81581

QY 181 CATAGGAACTCAATGCAATATATTAATTTGAGATTATAGAGAAGGAAATAGCAA 240
DB 81582 CATAGGAACTCAATGCAATATATTAATTTGAGATTATAGAGAAGGAAATAGCAA 81641

QY 241 ATGACACAATTAACAATGTGTGCGTGAGAGACATCTTTGAGGTCATGAGTT 300
DB 81642 ATGACACAATTAACAATGTGTGCGTGAGAGACATCTTTGAGGTCATGAGTT 81701

QY 301 TGTAGTTAACATCATATTTGTATATAGTGAACCTGACCAAAATATAGCAGCTT 360
DB 81702 TGTAGTTAACATCATATTTGTATATAGTGAACCTGACCAAAATATAGCAGCTT 81761

QY 361 GAACTGGCTTTCACATCTTGAATTTGACCAAGTGTCTTATATATGACATCTAAT 420
DB 81762 GAACTGGCTTTCACATCTTGAATTTGACCAAGTGTCTTATATATGACATCTAAT 81821

QY 421 GTAAATCCAGAACTTGACATCGTTAAATATATTAATGTGTACATTTCAATGTG 480
DB 81822 GTAAATCCAGAACTTGACATCGTTAAATATATTAATGTGTACATTTCAATGTG 81881

QY 481 GCATTAATATGCTTCACAGTAAATCTGAAAAACCTGATTTGATGAAAGCTGCTT 540
DB 81882 GCATTAATATGCTTCACAGTAAATCTGAAAAACCTGATTTGATGAAAGCTGCTT 81941

QY 541 TCAATTACTGAGTCTTGACATACATCTTTTATGAGTATGAATAAATCAATTTT 600
DB 81942 TCAATTACTGAGTCTTGACATACATCTTTTATGAGTATGAATAAATCAATTTT 82001

QY 601 AAACGTAAATCTTAATCTTGACATTTCAAAATTTCTTTCTTTCTTTCTTTT 660
DB 82002 AAACGTAAATCTTAATCTTGACATTTCAAAATTTCTTTCTTTCTTTCTTTT 82061

QY 661 TTTTTTTTGA 672
DB 82062 TTTTTTTTGA 82073

RESULT 3
AX197746/c 621 bp DNA linear PAT 29-AUG-2001
LOCUS
DEFINITION Sequence 201 from Patent W00151513.
ACCESSION AX197746
VERSION AX197746.1 GI:15388061
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS 1 (bases 1 to 621)
TITLE Algate, P.A.
JOURNAL Ovarian tumor-associated sequences
Patent: WO 0151513-A 201 19-JUL-2001;
CORIXA CORPORATION (US)
FEATURES
source 1..621
Location/Qualifiers
BASE COUNT 207 a 118 c 82 g 195 t 19 others
ORIGIN
Query Match 68.7%; Score 461.6; DB 6; Length 621;
Best Local Similarity 91.3%; Pred. No. 1.1e-66;
Matches 525; Conservative 0; Mismatches 44; Indels 6; Gaps 4;

QY 26 AATTAAGCTTATTAATGTGGGAAAAATTAATCTGTGTCCTTTAAATTCATTTTATT 85
DB 569 AATTAAGCTTATTAATGTGGGAAAAATTAATCTGTGTCCTTTAAATTCATTTTATT 513

QY 86 TAAATACATTTTCAGATGAAACAAAGATGAAATTTATTTGAATTTTCTGTGCT 145
DB 512 TAAAGCTATTTTCAGATGAAACAAAGATGAAATTTATTTGAATTTTCTGTGCT 455

QY 146 TTTTCTGTTTCAGATTAAGAGAGATGGTGATGATGATAGGAACTCAATGATAT 205
DB 454 TTTTCCGTTTCAGATTAAGAGAGAGATGGTGATGATGATGATGATGATGATGAT 396

QY 206 ATTAATTTGAAGATTATGAAGAAGGAAATAGCAATGACACAAATACAAATGTGTG 265
DB 395 ATTAATTTGAAGATTATGAAGAAGGAAATAGCAATGACACAAATACAAATGTGTG 336

QY 266 GCGGGGAGCAAGCATCTTGAAGGTCATGAGTTGTATGTTAATCATCATATTTGT 325
DB 335 GCGGGGAGCAAGCATCTTGAAGGTCATGAGTTGTATGTTAATCATCATATTTGT 276

QY 326 AATAGTGAACCTGTACTCAAAATATTAAGCAGCTTGAACTGCTTATACCAATCTGAAA 385
DB 275 AATAGTGAACCTGTACTCAAAATATTAAGCAGCTTGAACTGCTTATACCAATCTGAAA 216

QY 386 TTGGACCAAGAGTGTCTTATATATGAGATCTAATGTAAATCCAAACTTGGACCTCAT 445
DB 215 TTGGACCAAGAGTGTCTTATATATGAGATCTAATGTAAATCCAAACTTGGACCTCAT 156

QY 446 CGTTAAATATTAATGTATGTAACATTCAAATGTGTGATTAATATTCCTCCACAGTAA 505
DB 155 CGTTAAATATTAATGTATGTAACATTCAAATGTGTGATTAATATTCCTCCACAGTAA 96

QY 506 ATCTGAAGAACTGATTTGTATGAAAGCTGCTTTCTATTTACTTGAGCTTGATACATA 565
DB 95 ATCTGAAGAACTGATTTGTATGAAAGCTGCTTTCTATTTACTTGAGCTTGATACATA 36

QY 566 CATACTTTTTTATGAGCTATGAATAAATCAATTTT 600
DB 35 CATACTTTTTTATGAGCTATGAATAAATCAATTTT 1

RESULT 4
AX208351/c 621 bp DNA linear PAT 31-AUG-2001
LOCUS

DEFINITION	Sequence 191 from Patent WO0157207.
ACCESSION	AX208351
VERSION	AX208351.1 GI:15422774
KEYWORDS	
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS	Algate,P.A. and Mannion,J.
TITLE	Compositions and methods for the therapy and diagnosis of ovarian cancer
JOURNAL	Patent: WO 0157207-A 191 09-AUG-2001; CORIXA CORPORATION (US)
FEATURES	Location/Qualifiers
Source	1..621
BASE COUNT	207 a 118 c 82 g 195 t 19 others
ORIGIN	
Query Match	68.7%; Score 461.6; DB 6; Length 621;
Best Local Similarity	91.3%; Pred. No. 1,1e-66;
Matches 525; Conservative	0; Mismatches 44; Indels 6; Gaps 4;
OY	26 AATAAGCTCTATAATGTGGGAAAAATATCTGTCTCTTAATTTCATTTTATT 85
Db	569 AATAAGCTCTATAATGTGGGAAAAATTTCTGTG---TTCNTTMTTTCATTNTATT 513
OY	86 TAAATCAATTTTTCAGATGAAACAAAAGATTGAAAAATTTAGATTTTTCGTGCT 145
Db	512 TAAACCTATTTTTCAGATG-ACNAAAGATTGAAAAATTTATTTAG-ATTTTTCGTGCT 455
OY	146 TTTTCCCTTTTCAGTAAAGGATGGTGGATGATCATAGGAACTCATGATTAACAT 205
Db	454 TTTTCCNGTTTCAGTAAAGGAGGCGGAAATGACATTA-GGAACTCAATGATTAACAT 396
OY	206 ATAATTGAGATTTTATAGAGAGGAAATAGCAAAATGACACAATTCACAATGTGTGT 265
Db	395 ATACTTTGAAGATTTATTAAGAAAGGAAATAGCAAAATGCGACAATTTACNAAATGTGTGT 336
OY	266 GCGTGGAGCAAGAAATATCTTGAAGCTATAGTTTGTGTTAATCAATCATATTTGT 325
Db	335 GCGTGGAGCAAGACATCTTGAAGCTATAGTTTGTGTTAATCAATCATATTTGT 276
OY	336 AATAGTGAACCTGTACTCAAAAATATTAAGCAGCTTGAACATGCGCTTACCAATCTTGAA 385
Db	275 AATATGAAACCTGTACTCAAAAATATTAAGCAGCTTGAACATGCGCTTTCACATCTTGAA 216
OY	386 TTTGACCAACAAGTCTTATATATGACAGATCTAATGTAAATCCGAACCTGGACTCCAT 445
Db	215 TTTGACCAACAAGTCTTATATATGACAGATCTAATGTAAATCCGAACCTGGACTCCAT 156
OY	446 CGTTAAATTTATTTATGTGTAACTTCAAAATGTGCATTTAAATTTGCTCCACAGTAA 505
Db	155 CGTTAAATTTATTTATGTGTAACTTCAAAATGTGCATTTAAATTTGCTCCACAGTAA 96
OY	506 ATCTGAAAACTGATTTGTGATTTGAAGCTGCTTTGATTTTACTTGAGTCTTTGACATA 565
Db	95 ATCTGAAAAAGATTTGTGATTTGAAGCTGCTTTGATTTTACTTGAGTCTTTGACATA 36
OY	566 CATACTTTTATGAGCTATGAAATTAACATTTT 600
Db	35 CATGCTTTTATGAGCTATGAAATTAACATTTT 1
RESULT 5	
LOCUS	AX014895 1545 bp DNA linear PAT 07-SEP-2000
DEFINITION	Sequence 94 from Patent WO9955040.
ACCESSION	AX014895
VERSION	AX014895.1 GI:10041162
KEYWORDS	

SOURCE	human.
ORGANISM	Homo sapiens
TITLE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. (bases 1 to 1545)
REFERENCE	Schmitt,A., Speecht,T., Dahl,E., Hinemann,B., Rosenthal,A. and Piarlsky,C.
JOURNAL	Human nucleic acid sequences from ovarian lnmour tissue Patent: WO 9553040-A 94 21-OCT-1999.
FEATURES	SCHEMTT AMIN (DE); SPECHT THOMAS (DE); DAHL EDGAR (DE); HINZMANN BENND (DE); ROSENTHAL ANDRE (DE); METAGEN GES FUER GENOMFORSCHUN (DE); PIARLSKY CHRISTIAN (DE)
SOURCE	location/Qualifiers 1..1545
BASE COUNT	/organism="Homo sapiens" /db_xref="taxon:9606"
ORIGIN	460 a 307 c 369 g 409 t
Query Match	67.2%; Score 451.4; DB 6; Length 1545;
Best Local Similarity	99.8%; Pred. No. 4.1e-65;
Matches	452; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Dy	156 TCAGATAAAGGAGATGGGTGAGATCATAGGACACTCAATGCATACATATAAATTGA 215
Dd	1077 TCAGATAAAGGAGATGGGTGAGATCATAGGACACTCAATGCATACATATAAATTGA 1136
OY	216 GATTATAGAAGAGGAATATGCAATATGACACAATTTACAATGTGTGTCGGTGACG 275
Dd	1137 GATTATAGAAGAGGAATATGCAATATGACACAATTTACAATGTGTGTCGGTGACG 1196
OY	276 AAGACATCTTTTAAGAGCTCAGATTTGTTAGTTAACATCAATATATTTGTAATAGTGA 335
Dd	1197 AAGACATCTTTTAAGAGCTCAGATTTGTTAGTTAACATCAATATATTTGTAATAGTGA 1256
OY	336 CCTGTACTCAAATATATAGCAGCTTGAACCTGGCTTACCAATCTTGAATTTGACCACA 395
Dd	1257 CCTGTACTCAAATATATAGCAGCTTGAACCTGGCTTACCAATCTTGAATTTGACCACA 1316
OY	396 AGTGCCTTATATATGACAGTCTAATGTAAAAATCCAGAACTGGACTCCATCGTTAAAATT 455
Dd	1317 AGTGCCTTATATATGACAGTCTAATGTAAAAATCCAGAACTGGACTCCATCGTTAAAATT 1376
OY	456 ATTATATGTGTAAACATTTCAATGTGCAATTAAATATGCTTCCACAGTAAAAATCTGAAAA 515
Dd	1377 ATTATATGTGTAAACATTTCAATGTGCAATTAAATATGCTTCCACAGTAAAAATCTGAAAA 1436
OY	516 CTGATTTTGATGATGAAGAGCTGCCCTTTCATATTAAGTGAAGTCTGTACATACATCTTTT 575
Dd	1437 CTGATTTTGATGATGAAGAGCTGCCCTTTCATATTAAGTGAAGTCTGTACATACATCTTTT 1496
OY	576 TATGAGCTATGAAATATAAATCATTTTAACTGAA 608
Dd	1497 TATGAGCTATGAAATATAAATCATTTTAACTGAA 1529
RESULT 6	
AKO26585	1565 bp mRNA linear PRI 29-SEP-2000
LOCUS	Homo sapiens CDNA: FLJ22932.f1s, clone KAT07515, highly similar to HUNGAWA Human (clone GA733-2-2) carcinoma-associated antigen GA733-2-mRNA.
ACCESSION	AKO26585
VERSION	AKO26585.1 GI:10439469
KEYWORDS	oligo cloning; fts (full insert sequence).
SOURCE	Homo sapiens signal:ring cell carcinoma cell_line:KATO III CDNA to mRNA, clone.lib:KAT clone:KAT07515.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. (sites)
AUTHORS	Ota,T., Suzuki,Y., Obayashi,M., Nishi,T., Shibahara,T., Tanaka,T.,

TITLE Nakamura,Y., Isogai,T. and Sugano,S.
JOURNAL NEDO human cDNA sequencing project
REFERENCE Unpublished
AUTHORS 2 (bases 1 to 1565)
Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T., Isogai,T.,
Shibahara,T., Tanaka,T. and Nakamura,Y.
TITLE Direct Submission
JOURNAL Submitted (29-AUG-2000) Sumio Sugano, Institute of Medical Science,
University of Tokyo, Laboratory of Genome Structure Analysis, Human
Genome Center, Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639,
Japan (E-mail:cdna@ems.u-tokyo.ac.jp, Tel:81-3-5449-5286,
Fax:81-3-5449-5416)
COMMENT NEDO human cDNA sequencing project supported by Ministry of
International Trade and Industry of Japan: cDNA full insert
sequencing: Research Association for Biotechnology: cDNA library
construction, 5'- & 3'-end one pass sequencing: Department of
Virology and Human Genome Center, Institute of Medical Science,
University of Tokyo (partly supported by Science and Technology
Agency).
FEATURES
source Location/Qualifiers
1..1565
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="KAT07515"
/cell_line="KATO III"
/cell_type="signed-ring cell carcinoma"
/clone_lib="KAT"
/note="cloning vector pME18SFL3"
misc_feature
1..1565
/note="highly similar to HUMGA7A Human (clone GA733-2-2)
carcinoma-associated antigen GA733-2 mRNA"
BASE COUNT 468 a 321 c 367 g 409 t
ORIGIN
Query Match 66.9%; Score 449.8; DB 9; Length 1565;
Best Local Similarity 99.6%; Pred. No. 7.5e-65;
Matches 451; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 156 TCAGATTAAGAGAGATGGTGAGATGATAGGAACTCAATGCACTATATATATTGAA 215
1095 TGAGATAAAGAGATGGTGAGATGATAGGAACTCAATGCACTATATATATTGAA 1154
QY 216 GATTATAGAAGAGAGAAATAGCAATGGACACAAATTACAAATGTGTGCGTGGAGC 275
1155 GATTATAGAAGAGAGAAATAGCAATGGACACAAATTACAAATGTGTGCGTGGAGC 1214
QY 276 AAGACATCTTTGAAGGTCATGAGTTGTTAGTTAAATCATATATATTTGTAATAGTGA 335
1215 AAGACATCTTTGAAGGTCATGAGTTGTTAGTTAAATCATATATATTTGTAATAGTGA 1274
QY 336 CCGTACTCAAAATATATAGCAGCTTGAACCTGGCTTACCAATCTTGAATTTGACCACA 395
1275 CCGTACTCAAAATATATAGCAGCTTGAACCTGGCTTACCAATCTTGAATTTGACCACA 1334
QY 396 AGTGTCTTATATATGAGATCTAATGTAAATCCAGAACTGGAGCTCATGTTAAATTT 455
1335 AGTGTCTTATATATGAGATCTAATGTAAATCCAGAACTGGAGCTCATGTTAAATTT 1394
QY 456 ATTATGTGTGAACATTCATGTCATTAATATGCTTCCACAGTAAATGCTGAAAAA 515
1395 ATTATGTGTGAACATTCATGTCATTAATATGCTTCCACAGTAAATGCTGAAAAA 1454
QY 516 CCGATTGTCATGTAAGAGCTGCTTCTATTCTAGTCTGCTGTCATCATCTTTT 575
1455 CCGATTGTCATGTAAGAGCTGCTTCTATTCTAGTCTGCTGTCATCATCTTTT 1514
QY 576 TATGAGCTATGAATAAATCAATTTTAACTGAA 608
1515 TATGAGCTATGAATAAATCAATTTTAACTGAA 1547
RESULT 7
HSKSA

LOCUS HSKSA 1504 bp mRNA linear PRI 22-MAR-1995
DEFINITION Human mRNA for adenocarcinoma-associated antigen (KSA).
ACCESSION X14758
VERSION X14758.1 GI:34186
KEYWORDS antigen; cell surface glycoprotein.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 1504)
AUTHORS Strnad,J., Hamblton,A.E., Beavers,L.S., Gamboa,G.C., Apelgren,L.D.,
Taber,L.D., Sportsman,J.R., Bunni,T.F., Sharp,J.D. and Gadski,R.A.
Molecular cloning and characterization of a human
adenocarcinoma/epithelial cell surface antigen complementary DNA
Cancer Res. 49 (2), 314-317 (1989)
JOURNAL MEDLINE 89089570
PUBMED 2463074
FEATURES
source Location/Qualifiers
1..1504
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_line="UCLA-P3"
/clone_lib="lambda gtl1."
155..1099
CDS
/note="KSA preproantigen peptide"
/codon_start=1
/protein_id="CAA32870.1"
/db_xref="SWISS-PROT:P16422"
/db_xref="GI:34187"
/translation="MAPROYLAFGLLAATATPAADAECECYENKLAVNGFVNNR
OCQCTSVGQNTVCSKLAQKLVKAEKNGSKLQRKPBGLDNLNGLDPDDES
GLEKAKQCGSTWCNVNAGVRRTDKDEITGSESRVRYWIIELKIKARRPYDSK
SLRTALQKEITRYOLDPKFTFISLYENNVTIIRLVNVSQKTDQVDIADVAAYFEK
DYKGEISFHSKMDLVNGEOLDLPQGLIYIVDEKARPEFGMSQGLKAGLAVIVVV
MAVVGIVLYIVSRKKRMAKYKAEIKEMGENHRLN"
155..223
sig_peptide
/note="putative"
398..1096
mat_peptide
/product="mature KSA antigen (AA 59-291)"
224..1096
misc_feature
/note="KSA proantigen (AA 1-291)"
1504
polya_site
/note="polya site"
BASE COUNT 442 a 302 c 356 g 404 t
ORIGIN
Query Match 66.9%; Score 449.4; DB 9; Length 1504;
Best Local Similarity 99.8%; Pred. No. 8.8e-65;
Matches 450; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 156 TCAGATTAAGAGAGATGGTGAGATGATAGGAACTCAATGCACTATATATTGAA 215
1054 TGAGATAAAGAGATGGTGAGATGATAGGAACTCAATGCACTATATATTGAA 1113
QY 216 GATTATAGAAGAGAGAAATAGCAATGGACACAAATTACAAATGTGTGCGTGGAGC 275
1054 GATTATAGAAGAGAGAAATAGCAATGGACACAAATTACAAATGTGTGCGTGGAGC 1173
QY 276 AAGACATCTTTGAAGGTCATGAGTTGTTAGTTAAATCATATATTTGTAATAGTGA 335
1174 AAGACATCTTTGAAGGTCATGAGTTGTTAGTTAAATCATATATTTGTAATAGTGA 1233
QY 336 CCGTACTCAAAATATATAGCAGCTTGAACCTGGCTTACCAATCTTGAATTTGACCACA 395
1234 CCGTACTCAAAATATATAGCAGCTTGAACCTGGCTTACCAATCTTGAATTTGACCACA 1293
QY 396 AGTGTCTTATATATGAGATCTAATGTAAATCCAGAACTGGAGCTCATGTTAAATTT 455
1294 AGTGTCTTATATATGAGATCTAATGTAAATCCAGAACTGGAGCTCATGTTAAATTT 1353
QY 456 ATTATGTGTGAACATTCATGTCATTAATATGCTTCCACAGTAAATGCTGAAAAA 515
1354 ATTATGTGTGAACATTCATGTCATTAATATGCTTCCACAGTAAATGCTGAAAAA 1413
DB

```

OY 516 CTGATTTCGATGGAAGCGCTTCTTATTTACTGTGAGCTGTGACATACACTTTT 575
Db 1414 CTGATTTCGATGGAAGCGCTTCTTATTTACTGTGAGCTGTGACATACACTTTT 1473
OY 576 TATGAGCTATGAAATTAACATTTTAACTG 606
Db 1474 TATGAGCTATGAAATTAACATTTTAACTG 1504

RESULT 8
HUMKSA 1504 bp mRNA linear PRI 11-JUN-1993
LOCUS Human adenocarcinoma-associated antigen (KSA) mRNA, complete cds.
ACCESSION M32325
VERSION M32325.1 GI:186775
KEYWORDS adenocarcinoma-associated antigen.
SOURCE Human cell line UCLA-P3, cDNA to mRNA, clone AG[1,1338,933].
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 1504)
AUTHORS Strnad,J., Hamilton,A.E., Beavers,L.S., Gamboa,G.C., Apelgren,L.D.,
Taber,L.D., Sportsman,J.R., Bumol,T.F., Sharp,J.D. and Gadski,R.A.
TITLE Molecular cloning and characterization of a human
adenocarcinoma/epithelial cell surface antigen complementary DNA
JOURNAL Cancer Res. 49 (2), 314-317 (1989)
MEDLINE 89089570
PUBMED 2463074
COMMENT Draft entry and computer-readable sequence for [1] kindly submitted
by J.D. Sharp, 22-FEB-1990.
There are a few base differences between the sequence presented
here and that which appears in entry M26481. The difference occurs
mostly in the Poly-A signal.
FEATURES
Source
Location/Qualifiers
1..1504
/organism="Homo sapiens"
/db_xref="taxon:9606"
<1..1504
/product="adenocarcinoma-associated antigen mRNA"
155..1099
/feature="adenocarcinoma-associated antigen precursor (KSA)"
/db_xref="GI:307091"
/codon_start=1
/protein_id="AA36151.1"
/translation="MAPPVLAFLGLLAATATPAAAOECVNCYLAIVNCPVNNR
OCQCTSVGAONTVCISLAKCLVMKEMNGSKGRARKEGALQNDGLYDPCDSS
GLFKAKCNGSTICWCYNTAGVTRTDKDTETCSERVRTWIIILKHAKREYDSK
SLRTALOKETITTRQIDPKFTTSILYENNVITIDLVONSSOKTQNDVIDIADVAIFPK
DVKGESLFHKKMDLVNGEOLDIDPCOTLIYVYDEKAPESMGLAKAGYIAIVVVV
MAYVAGIVLVISIKKRMATYERKEIREKEMHRELNA"
155..223
/sig_peptide
/feature="adenocarcinoma-associated antigen signal peptide"
mat_peptide 398..1096
/product="adenocarcinoma-associated antigen"
polya_signal 1486..1491
BASE COUNT 442 a 302 c 356 g 404 t
ORIGIN
Query Match 66.98; Score 449.4; DB 9; Length 1504;
Best Local Similarity 99.88; Pred. No. 8.8e-65;
Matches 450; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 156 TCACATTAAGGAGTGGTGCATGATGAGCACTCAATACATATTAATTGAA 215
Db 1054 TGACATTAAGGAGTGGTGCATGATGAGCACTCAATACATATTAATTGAA 1113
OY 216 GATTATAGAGAAAGCAATAGCAATGACACAAATTAAGTGTGCTGCGAGCG 275
Db 1114 GATTATAGAGAAAGCAATAGCAATGACACAAATTAAGTGTGCTGCGAGCG 1173
OY 276 AAGCATCTTTGAAGTGCATGTTGTTAATTAACTATATTTGTAATAGTGAA 335
Db 1174 AAGCATCTTTGAAGTGCATGTTGTTAATTAACTATATTTGTAATAGTGAA 1233
OY 336 CCTGTACTCAAAAATATTAAGCAGCTTGAACCTGGCTTTACCAATCTTGAATTTGACACA 395
Db 1234 CCTGTACTCAAAAATATTAAGCAGCTTGAACCTGGCTTTACCAATCTTGAATTTGACACA 1293
OY 396 AGTGCTTATATATGACATCTAATAGTAATCCGAACCTTGACATCGTTAAATTT 455
Db 1294 AGTGCTTATATATGACATCTAATAGTAATCCGAACCTTGACATCGTTAAATTT 1353
OY 456 ATTATGTGTAACATTCGAATGTGCAATTAATATGCTCCACAGTAATCTGAAAA 515
Db 1354 ATTATGTGTAACATTCGAATGTGCAATTAATATGCTCCACAGTAATCTGAAAA 1413
OY 516 CTGATTTCGATGGAAGCGCTTCTTATTTACTGTGAGCTGTGACATACACTTTT 575
Db 1414 CTGATTTCGATGGAAGCGCTTCTTATTTACTGTGAGCTGTGACATACACTTTT 1473
OY 576 TATGAGCTATGAAATTAACATTTTAACTG 606
Db 1474 TATGAGCTATGAAATTAACATTTTAACTG 1504

RESULT 9
HUMGA7A 1457 bp mRNA linear PRI 27-APR-1993
LOCUS Human (clone GA733-2-2) carcinoma-associated antigen GA733-2 mRNA,
complete cds.
ACCESSION M33011
VERSION M33011.1 GI:182895
KEYWORDS carcinoma-associated antigen GA733-2.
SOURCE Homo sapiens cDNA to mRNA.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 1457)
AUTHORS Szala,S., Froehlich,M., Scollon,M., Kasai,Y., Steplewski,L.,
Koprowski,H. and Linenbach,A.J.
TITLE Molecular cloning of cDNA for the carcinoma-associated antigen
GA733-2
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 87 (9), 3542-3546 (1990)
MEDLINE 90239051
PUBMED 2333300
COMMENT Draft entry and computer-readable sequence for [1] kindly submitted
by A.J. Linenbach, 22-MAR-1990.
FEATURES
Source
Location/Qualifiers
1..1457
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_line="SW948"
/cell_type="colorectal carcinoma"
<1..1457
/rna
108..1052
/codon_start=1
/product="carcinoma-associated antigen GA733-2"
/protein_id="AA35861.1"
/translation="MAPPVLAFLGLLAATATPAAAOECVNCYLAIVNCPVNNR
OCQCTSVGAONTVCISLAKCLVMKEMNGSKGRARKEGALQNDGLYDPCDSS
GLFKAKCNGSTICWCYNTAGVTRTDKDTETCSERVRTWIIILKHAKREYDSK
SLRTALOKETITTRQIDPKFTTSILYENNVITIDLVONSSOKTQNDVIDIADVAIFPK
DVKGESLFHKKMDLVNGEOLDIDPCOTLIYVYDEKAPESMGLAKAGYIAIVVVV
IAYVAGIVLVISIKKRMATYERKEIREKEMHRELNA"
108..155
/sig_peptide
/feature="putative"
mat_peptide 156..1049
/product="carcinoma-associated antigen GA733-2"
polya_signal 1439..1444
BASE COUNT 437 a 285 c 339 g 396 t
ORIGIN
Query Match 66.68; Score 447.8; DB 9; Length 1457;
Best Local Similarity 99.68; Pred. No. 1.6e-64;

```


Matches	4.9:	Conservative	0:	Mismatches	2:	Indels	0:	Gaps	0:
QY	156	TCAGATAAAGAGATGGGTGAGATGCATAGGGAGACTCAATGCATATATATTAATTGAA	215						
Db	1007	TGAGATTAAGGAGATGGGTGAGATGCATAGGGAGACTCAATGCATATATATTAATTGAA	1066						
QY	216	GATTATAGAGAAGGAATAGCAATAGGACACCAATTAACAATGTGTGGCTGGGAGC	275						
Db	1067	GATTATAGAGAAGGAATAGCAATAGGACACCAATTAACAATGTGTGTGGCTGGGAGC	1126						
QY	276	AAGACATCTTGAAGCGTCATGAGTTTGTACTTTAACATCATATATTTGTAACTGAAA	335						
Db	1127	AAGACATCTTGAAGCGTCATGAGTTTGTACTTTAACATCATATCTGTATAGTAAA	1186						
QY	336	CCTGTACACCAAAATTTAAAGCAGCTTGAAGACGGCTTTACCATCTTGAATTGGACACA	395						
Db	1187	CCTGTACACCAAAATTTAAAGCAGCTTGAAGACGGCTTTACCAATCTTGAATTGGACACA	1246						
QY	396	AGTGTCTTATATATGTCAGATCTTAATGATAAAATCCAGAACTTGGACTCCATCGTTAAATT	455						
Db	1247	AGTGTCTTATATATGTCAGATCTTAATGATAAAATCCAGAACTTGGACTCCATCGTTAAATT	1306						
QY	456	ATTATATGCTGAACATTCCAATGTGTGCATTAATATGCTTCCACAGTAAATCTGAAAA	515						
Db	1307	ATTATATGCTGAACATTCCAATGTGTGCATTAATATGCTTCCACAGTAAATCTGAAAA	1366						
QY	516	CTGATTTTGCATGTAAGCTGCCCTTCTATTTACTTGAAGTCTTGTACATACATACTTTT	575						
Db	1367	CTGATTTTGCATGTAAGCTGCCCTTCTATTTACTTGAAGTCTTGTACATACATACTTTT	1426						
QY	576	TATGAGCTATGAATATAAAACATTTTAAACTG	606						
Db	1427	TATGAGCTATGAATATAAAACATTTTAAACTG	1457						

LOCUS	AX395799/c	637 bp	DNA	linear	PAT 18-MAY-2002
DEFINITION	Sequence 14 from Patent WO0212328.				
ACCESSION	AX395799				
VERSION	AX395799.1	GI:21066546			
KEYWORDS					
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
TITLE	1 King,G.E., Meagher,M.J., Xu,J. and Secrist,H. Compositions and methods for the therapy and diagnosis of colon cancer				
JOURNAL	Patent: WO 0212328-A 14 14-FEB-2002;				
FEATURES	CORIXA CORPORATION (US) Location/Qualifiers				
source	1..637				
	/organism="Homo sapiens"				
	/db_xref="taxon:9606"				
BASE COUNT	205 a 136 c 84 g 212 t				
ORIGIN					
Query Match	66.0%; Score 443.4; DB 6; Length 637;				
Best Local Similarity	99.8%; Pred. NO. 1e-63;				
Matches 444: Conservative	0; Mismatches 1; Indels 0; Gaps 0				
Oy	156 TCAGATTAAGGAGATGGCGTGCAGATGCATGGGAAACATGCATGATATTAATTGGAA	215			
Db	445 TGAGATTAAGGAGATGGGAGATGCATATGGGAACCATGCATATCTATATATTTGAA	386			
Oy	216 GATTATAGAGAAGGGAATATACCAATGGACACAAATTTCAATATGTGTGGCTGGAGC	275			
Db	385 GATTATAGAGAAGGGAATATACCAATGGACACAAATTTCAATATGTGTGGCTGGAGC	326			
Oy	276 AAGACATCTTGAAGGTATGAGATTGTTAGTTTAACATCATATATTTGTAATAGTGA	335			

Db	325	AAGACATCTTGAAGGCGTCAGAGTTGGTTAGTTAAACATCATATATTTGTAATAGAGAA	266
Qy	336	CCTGTACTCAAAATATATAGCAGCGCTGAAACTGCGTTTACCAATCTGAAATTTGACCACA	395
Db	265	CCGTGTACTCAAAATATATAGCAGCGCTGAAACGCGCTTTACCAATCTGAAATTTGACCACA	206
Qy	396	AGTGTCTATATATATGCGAGATCTATATGTAAATCCAGAACTTGCACTCCATCGTTAAATTT	455
Db	205	AGTGTCTATATATATGCGAGATCTATATGTAAATCCAGAACTTGCACTCCATCGTTAAATTT	146
Qy	456	ATTATATGTGTAACTTCAAAATGTCGATTAATAATGCTCCACAGTAATAATCTGAAAA	515
Db	145	ATTATATGTGTAACTTCAAAATGTCGATTAATAATGCTCCACAGTAATAATCTGAAAA	86
Qy	516	CTGATTTGTGATTAAGAGCTGCCCTTTCTATTTACTTACTAGTCTTGACATACATACTTTT	575
Db	85	CTGATTTGTGATTAAGAGCTGCCCTTTCTATTTACTTACTAGTCTTGACATACATACTTTT	26
Qy	576	TATGAGCTATGAATTAATAACATTTT	600
Db	25	TATGAGCTATGAATTAATAACATTTT	1

AX397344/c	AX397344	481 bp	DNA	linear	PAT 18-MAY-2002
LOCUS	Sequence	1559 from Patent WO0212328.			
DEFINITION	AX397344				
ACCESSION	AX397344				
VERSION	AX397344..1	GI:21068091			
KEYWORDS					
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Gracilata; Vertebrata; Euteleostomi;				
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.				
TITLE	1 King, G.E., Meagher, M.J., Xu, J. and Secrist, H.				
JOURNAL	Compositions and methods for the therapy and diagnosis of colon cancer				
FEATURES	Patent: WO 0212328-A 1559 14-FEB-2002;				
source	CORIXA CORPORATION (US)				
	Location/Qualifiers				
	1..481				
	/organism="Homo sapiens"				
	/db_xref="taxon:9606"				
BASE COUNT	149 a 93 c 64 g 174 t				
ORIGIN	1 others				
Query Match	65.8%; Score 442.4; DB 6; Length 481;				
Best Local Similarity	99.6%; Pred. No. 1.6e-63;				
Matches 443; Conservative	0; Mismatches 2; Indels 0; Gaps 0;				
QY	156 TCAGATMAAGAGAGATGGGTGAGATGCATAGGAGACCTCAATGCATACATATATAATTTGAA	215			
DB	445 TCAGATMAAGAGAGATGGGTGAGATGCATAGGAGACCTCAATGCATACATATATAATTTGAA	386			
QY	216 GATTATTAGAGAAGGGAATAGCAAAATGACACCAATTAACAATCTGTGTGCGTGGGACG	275			
DB	385 GATTATTAGAGAAGGGAATAGCAAAATGACACCAATTAACAATCTGTGTGCGTGGGACG	326			
QY	276 AAGACATCTTGAAGGTCATGAGTTGTAGTTAAATCATATATTTGTATAGTGAA	335			
DB	325 AAGACATCTTGAAGGTCATGAGTTGTAGTTAAATCATATATTTGTATAGTGAA	266			
QY	336 CCTGTACTCAAAATATAAGCAGCTTGAACCTGGCTTTACCAATCTTGAATTTGACACA	395			
DB	265 CCTGTACTCAAAATATAAGCAGCTTGAACCTGGCTTTACCAATCTTGAATTTGACACA	206			
QY	396 AGTGTCTTATATATGACATCTAATGTAAATTCACAACTGGAGCTCCATGCTTAAATT	455			
DB	205 AGTGTCTTATATATGACATCTAATGTAAATTCACAACTGGAGCTCCATGCTTAAATT	146			
QY	456 ATTATATGTATACATTCAATGTGTGTCATTAATATGCTCCACAGTAAATCTGAAAA	515			

```

Db 145 ATTATGTGTACATCAATGATGTGTCATTAATATGCTCCACAGTAAATCTGAAAAA 86
Oy 516 CTGATTTGTGATGAAGCTGCTTCTATTTACTGATGCTGTGACATACACTTTT 575
Db 85 CTGATTTGTGATGAAGCTGCTTCTATTTACTGATGCTGTGACATACACTTTT 26
Oy 576 TATGAGCTATGAATTAATCAATTTT 600
Db 25 TATGAGCTATGAATTAATCAATTTT 1

RESULT 12
AX198887 545 bp DNA linear PAT 29-AUG-2001
LOCUS AX198887/c
DEFINITION Sequence 1342 from Patent WO0151513.
ACCESSION AX198887
VERSION AX198887.1 GI:15389213
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 545)
AUTHORS Algate,P.A.
TITLE Ovarian tumor-associated sequences
JOURNAL Patent: WO 0151513-A 1342 19-JUL-2001;
CORIXA CORPORATION (US)
FEATURES
Source location/Qualifiers
1..545
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 163 a 112 c 89 g 177 t 4 others
ORIGIN
Query Match 65.8%; Score 442.2; DB 6; Length 545;
Best Local Similarity 96.0%; Pred. No. 1.7e-63;
Matches 453; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

Oy 129 GAATTTTTCGTCGCTTTTCCCTGTTCCAGTAAAGAGATGGGTGATGATAGGGA 188
Db 472 GGAATTCGCCCTTACCGTGGTGGCGCGCGAGGTAAAGGATGGGTGATGATAGGGA 413
Oy 189 ACTCAATGCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 248
Db 412 ACTCAATGCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 353
Oy 249 AAATTACAATGTGTGCGCGTGGGAGCAAGACATCTTTGAAGGTCATGATTTTACTT 308
Db 352 AAATTACAATGTGTGCGCGTGGGAGCAAGACATCTTTGAAGGTCATGATTTTACTT 293
Oy 309 TAACATCTATATTTGTATATAGTGAACCTGACTCAAAATTAAGCAGCTTGAACCTGG 368
Db 292 TAACATCTATATTTGTATATAGTGAACCTGACTCAAAATTAAGCAGCTTGAACCTGG 233
Oy 369 CTTTACCATCTTGAATTTTGAACAGAGTGTCTATATATGACAGATCAATGTAATATC 428
Db 232 CTTTACCATCTTGAATTTTGAACAGAGTGTCTATATATGACAGATCAATGTAATATC 173
Oy 429 CAGAAGCTTGACCTCATGTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTA 488
Db 112 CAGAAGCTTGACCTCATGTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTA 113
Oy 489 TATGCTTCCACAGTAAATCTGAAAAAGTATTTGATGTAAGTGGCTTTCTATTTA 548
Db 112 TATGCTTCCACAGTAAATCTGAAAAAGTATTTGATGTAAGTGGCTTTCTATTTA 53
Oy 549 CTTGAGCTTGTGACATACACTTTTATGAGCTATGAAAAATTAACATTTT 600
Db 52 CTTGAGCTTGTGACATACACTTTTATGAGCTATGAAAAATTAACATTTT 1

RESULT 13
HUMMAKSL4

```

```

LOCUS HUMMAKSL4 1511 bp mRNA linear PRI 09-JAN-1995
DEFINITION Homo sapiens KSI/4 antigen mRNA, complete cds.
ACCESSION M26481
VERSION M26481.1 GI:619789
KEYWORDS KSI/4 antigen.
SOURCE Homo sapiens lung adenocarcinoma CDNA to mRNA.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 1511)
AUTHORS Perez,M.S. and Walker,L.E.
TITLE Isolation and characterization of a cDNA encoding the KSI/4
JOURNAL J. Immunol. 142 (10), 3662-3667 (1989)
MEDLINE 89235177
PUBMED 2469722
COMMENT On Jan 10, 1995 this sequence version replaced gi:341508.
FEATURES
Source location/Qualifiers
1..1511
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_line="UCLA-P3"
/cell_type="B lymphoblast"
/tissue_type="lung adenocarcinoma"
160..1104
/codon_start=1
/product="KSI/4 antigen"
/protein_id="AA59543.1"
/db_xref="GI:619790"
/translation="MAPPOVLAIFGLLAATATATVAAOEEVCENYKLVANVCVNNNR
QOCCTVGAONTVICSKLAAKLVKAKEMNSKIGRRKPGGALONNDGLDPDDES
GLFRKQONGTSTGCVNTAGVRDKDTEITGSESVFTYIIIEIKRAREKPYDSK
SLRTALOEITRTROLDPEKFTITSLYENNVITIDVONSOKTQNDVDIADVAYFEK
DYKESLPHSKKMLDTVNGEOLDIDPOTLLIYYDERKPEESMCKRAGVLAIVVVV
MAVVAIGVIVLVISRRKRAKIEKRIKEMGHEHLNA"
BASE COUNT 443 a 305 c 359 g 404 t
ORIGIN
Query Match 65.4%; Score 439.4; DB 9; Length 1511;
Best Local Similarity 99.6%; Pred. No. 3.9e-63;
Matches 451; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

Oy 156 TCAGATTAAGGATGGGATGATGCATAGGAGACATCAATGATTAATTAATTTGAA 215
Db 1055 TGAGATTAAGGATGGGATGATGCATAGGAGACATCAATGATTAATTAATTTGAA 1118
Oy 216 GATTATAGAAGAGGAATATAGCAAAATGAGACAAATTAATTAATTAATTAATTA 275
Db 1119 GATTATAGAAGAGGAATATAGCAAAATGAGACAAATTAATTAATTAATTAATTA 1178
Oy 276 AAGACATCTTTAAGGTCTAGAGTTTGTATTAATCAATCAATATTTTATAGTAA 335
Db 1179 AAGACATCTTTAAGGTCTAGAGTTTGTATTAATCAATCAATATTTTATAGTAA 1238
Oy 336 CCTGTACTCAAAATTAATTAAGCAGCTTGAACCTGCTTACCAATCTGAAATTTGACACA 395
Db 1239 CCTGTACTCAAAATTAATTAAGCAGCTTGAACCTGCTTACCAATCTGAAATTTGACACA 1298
Oy 396 AGTGTCTTATATATGAGATCTAATGTAAATCCAGAACCTTGACATCTGTTAAAT 455
Db 1299 AGTGTCTTATATATGAGATCTAATGTAAATCCAGAACCTTGACATCTGTTAAAT 1358
Oy 456 ATTATGTGTACATTTCAATGTGTGATTAATATATGCTTCCACAGTAAATCTGAAAA 515
Db 1359 ATTATGTGTACATTTCAATGTGTGATTAATATATGCTTCCACAGTAAATCTGAAAA 1418
Oy 516 CTGATTTGTGATGAAGCTGCTTCTATTTACTGAGCTGTGACATACACTTTT 575

```

```

Db 1419 CTCATTGTCATTGAAAGCGCTTCTATTACTTGAGCTTGTACATACATCTTTT 1478
QY 576 TATGACCTATGAATAAATCAATTTTAAACTGAA 608
Db 1479 TATGACCTATG-AATTAACATTTTAACTGAA 1510

RESULT 14
AX209414/C 439 bp DNA linear PAT 31-AUG-2001
LOCUS AX209414
DEFINITION Sequence 1254 from Patent WO0157207.
ACCESSION AX209414
VERSION AX209414.1 GI:15423837
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 439)
AUTHORS Algate, P.A. and Mannion, J.
TITLE Compositions and methods for the therapy and diagnosis of ovarian
cancer
JOURNAL Patent: WO 0157207-A 1254 09-AUG-2001;
CORIXA CORPORATION (US)
FEATURES
source 1. 439
Location/Qualifiers
BASE COUNT 142 a 80 c 60 g 157 t
ORIGIN
Query Match 65.3%; Score 439; DB 6; Length 439;
Best Local Similarity 100.0%; Pred. No. 6e-63;
Matches 439; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 162 AAGAGATGGGTGAGATGATAGGAGACTGATGATATATATATATATATATAT 221
Db 439 AAGAGATGGGTGAGATGATAGGAGACTGATGATATATATATATATATATAT 380
QY 222 AGAGAGGGAATATAGCAATATGACACAAATTCATATGCTGGTGGAGAGACA 281
Db 379 AGAGAGGGAATATAGCAATATGACACAAATTCATATGCTGGTGGAGAGACA 320
QY 282 TCTTTGAAGTCTAGATTTAGTTAATCATATATATATATATATATATATATAT 341
Db 319 TCTTTGAAGTCTAGATTTAGTTAATCATATATATATATATATATATATATAT 260
QY 342 CTCAAATATAGCAGCTTGAACCTGCTTACCAATCTTGAATTTGACCAAGTGT 401
Db 259 CTCAAATATAGCAGCTTGAACCTGCTTACCAATCTTGAATTTGACCAAGTGT 200
QY 402 TTATATATGAGATCTAATGTAATCCGAATCTGAGCTCCATGCTTAAATATAT 461
Db 199 TTATATATGAGATCTAATGTAATCCGAATCTGAGCTCCATGCTTAAATATAT 140
QY 462 GGTCAACATTCAAATGTCATTAATATGCTTCCACAGTAATAATCTGAAAACTGAT 521
Db 139 GGTCAACATTCAAATGTCATTAATATGCTTCCACAGTAATAATCTGAAAACTGAT 80
QY 522 TGTGATGTAAGAGCTGCTTTCTATTTAGTGTGATGATACATACATCTTTTATGAG 561
Db 79 TGTGATGTAAGAGCTGCTTTCTATTTAGTGTGATGATACATACATCTTTTATGAG 20
QY 582 CTATGAATATAAATCATTTT 600
Db 19 CTATGAATATAAATCATTTT 1

RESULT 15
HUMCSAE 1372 bp mRNA linear PRI 27-APR-1993
LOCUS HUMCSAE
DEFINITION Human epithelial glycoprotein (EGP) mRNA, complete cds.

```

```

ACCESSION M32306
VERSION M32306.1 GI:181132
KEYWORDS cell surface antigen; cell surface glycoprotein; epithelial
glycoprotein; transmembrane protein; tumor-associated antigen.
SOURCE Human colon carcinoma cell line HT-29, cDNA to mRNA, (library
HT-29/cdms).
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 1372)
AUTHORS Simon, B., Podolsky, D. K., Moldenhauer, G., Isselbacher, K. J.,
Gallon, C., S. and Brand, S. J.
TITLE Epithelial glycoprotein is a member of a family of epithelial cell
surface antigens homologous to nidogen, a matrix adhesion protein
Proc. Natl. Acad. Sci. U.S.A. 87 (7), 2755-2759 (1990)
JOURNAL 90207276
MEDLINE 2108441
PUBMED
COMMENT Draft entry and computer-readable sequence (1) kindly submitted by
B. Simon, 23-FEB-1990.
FEATURES
source 1. 1372
Location/Qualifiers
CDS mRNA
/db_xref="taxon:9606"
<1. 1372
/product="EGP mRNA"
22. 966
/feature="epithelial glycoprotein (EGP) precursor"
/note="start-1
/protein_id="AAA35723.1"
/db_xref="GI:181133"
/translation="MAPPOLVLAFLGLLAATATATFAAOECCVENCYKLAIVNPFVNNR
OCQCTSVGADNTVICSKLAKLAELKAEKNGSKLGRARPEALONNDLTPDDDS
GLFPAKOCNGTSMCMCVNTRAGRTDKEITSEVRYWYIIEIKHARRKRPDSE
SLRHALQETITRTDKEITSLRENNVITIDLVONSQKTDNDIADVADYFEK
DKVGSLEFSKMDITVDNGEODLDDEGDTLYVDKRAEFESMGLKAGVIAVIVV
IADVAGIIVLVISRRKRAKRAKAEKEKEMHRELA"
22. 84
/note="epithelial glycoprotein signal peptide"
/mat_peptide 85. 963
/product="epithelial glycoprotein"
BASE COUNT 436 a 233 c 317 g 386 t
ORIGIN
Query Match 65.2%; Score 438.4; DB 5; Length 1372;
Best Local Similarity 98.7%; Pred. No. 5.8e-63;
Matches 442; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 156 TCAGATTAAGAGATGGGTGAGATGATAGGAGACTCAATGCAVAACTATATATTTGAA 215
Db 921 TGAGATTAAGAGATGGGTGAGATGATAGGAGACTCAATGCAVAACTATATATTTGAA 980
QY 216 GATTATGAAGAAGGGAATATAGCAATATGCAATATGCTGCTGGAGAGC 275
Db 981 GATTATGAAGAAGGGAATATAGCAATATGCAATATGCTGCTGGAGAGC 1040
QY 276 AAGACATCTTTGAAGTCTAGATTTAGTTAATCATATATATTTGTAATATGTAAGAA 335
Db 1041 AAGACATCTTTGAAGTCTAGATTTAGTTAATCATATATATTTGTAATATGTAAGAA 1100
QY 336 CCTGTACTCAAAATATATAGCAGCTTGAACCTGCTTTACCAATCTTGAATTTGACCACA 395
Db 1101 CCTGTACTCAAAATATATAGCAGCTTGAACCTGCTTTACCAATCTTGAATTTGACCACA 1160
QY 396 AGTGTCTTATATATGAGATCTAATGTAATAATGCAACAGCTGAGCCATCGTTAAATTT 455
Db 1161 AGTGTCTTATATATGAGATCTAATGTAATAATGCAACAGCTGAGCCATCGTTAAATTT 1220
QY 456 ATTATATGTAACATCAATATGTCATTAATATATATGCTCCACAGTAAATCTGAAAAA 515
Db 1221 ATTATATGTAACATCAATATGTCATTAATATATATGCTCCACAGTAAATCTGAAAAA 1280
QY 516 CTGATTTGATGTAAGAGCTGCTTTCTATTTAGTGTGATGATACATACATCTTTT 575

```

Db 1281 CTGATTGTGATTGAAGCTGCTTCTATTACTTGAGCTCTGTACATACATCTTTT 1340
Oy 576 TATGAGCTATGAATATAACATTTTAA 603
|||||
Db 1341 TATGAGCTATGAATATAACATTTTAA 1368
|||||

Search completed: November 26, 2002, 22:48:44
Job time : 1288.24 secs

THIS PAGE BLANK (USPTO)

OY	206	ATFTTTGAACTATTATGAAGAAGCGAATAATAGCAAATGACACCAAAATACAATNGTGT	265
Db	395	ATACCTTGAAATATTATGAAAGAAAAGGGAAAAATAGCAAAATGGCACAAAATTAACMAATGTGT	336
OY	266	GCGTGGACGAAAGACATCTTTGAAGGCATAGAGTTGTAGTTAACATCATATATTTGT	325
Db	335	GCGTGGGACGAAGACATCTTTGAAGGCATAGAGTTGTAGTTAACATCATATATTTGT	276
OY	326	AATAGTGAACCTGTACTCAAAATATATPAAGCAGCTTGAACCTGGCTTTACCAAATCTGAAA	385
Db	275	AATAGTGAANCCCTGTACTCAAAATATATPAAGCAGCTTGANCTGGCTTTNCCAATCTTGAAA	216
OY	386	TTTACCAACAAGTGTCTTATATATGACGAGTCATATGTAATAATCCAGACACTGGACTGCCAT	445
Db	215	TTTTACCAACAAGTGTCTTATATATGACGAGTCATATGTAATAATCCAGAGCTTGGACTGCCAT	156
OY	446	CGTTAAATATTTATATGTAACATTTAAATAGTGTGATTAATATATGCTCCACAGATAAA	505
Db	155	CGTTAAATATTTATANGTGTACCTTTAAATGTGTGATTAATATATGCTCCACAGATAAA	96
OY	506	ATCTGAAAAAAGCTGATTTGTGATTTGAAGAGTGCCTTTCTATTATCTAGACTTGTACATA	565
Db	95	ATCTGAAAAAAGCTGATTTGTGATTTGAAGAGTGCCTTTCTATTATCTAGACTTGTACATA	36
OY	566	CATACCTTTTATATGAGCTATGAAATATAACAATTTT	600
Db	35	CATGCTTTTATATGAGCTATGAAATATAACAATTTT	1

XX	AAH82577/c	AAH82577 standard; cDNA; 621 BP.
XX	AAH82577:	
XX	25-SEP-2001	(first entry)
XX	Human ovarian tumour associated polynucleotide sequence SEQ ID NO:201.	
XX	Human; ovarian tumour; ovarian cancer; diagnosis; gene therapy;	
XX	immunogenic; vaccine; ss.	
XX	Homo sapiens.	
XX	WO200151513-A2.	
XX	19-JUL-2001.	
XX	16-JAN-2001; 2001WO-US01575.	
XX	14-JAN-2000; 2000US-0176722.	
XX	(CORI-) CORIXA CORP.	
XX	Algate PA:	
XX	WPI; 2001-425866/45.	
XX	Novel ovarian tumor proteins, and nucleic acids encoding them, used to treat and diagnose cancers, particularly ovarian cancer -	
XX	Claim 5; Page 99; 338bp; English.	
XX	AAH82377 to AAH83878 represent human ovarian tumour-associated polynucleotide sequences which encode ovarian tumour proteins. The ovarian tumour protein and polynucleotide sequences have cytostatic activity, and can be used in gene therapy and vaccine production. The ovarian tumour proteins and polynucleotides can be used to inhibit the development of cancer, particularly ovarian cancer. They can also be used to diagnose the onset and progression of cancer.	
XX	Sequence 621 BP; 207 A; 118 C; 82 G; 195 T; 19 other;	

Query Match	68.7%	Score 461.6	DB 22	Length 621
Best Local Similarity	91.3%	Pred. No. 2.8e-71		
Matches 525	Conservative 0	Mismatches 44	Indels 6	Gaps 4
OY	AAATTAAGCTCTTTAAATGAGGGAATAAATATATCTGTGTCCTTAAATTCATTTTAAAT	85		
Db	569 AAATAGAGTNTTAAATGGGGGAAAAATTTCTGTG---TTCNTTNATTTCAATTTNTAT	513		
OY	86 TAATACATATTTTACAGATGAACAAAAAGATTGAAAAATATATTTAGATTTTTTCTGTGCT	145		
Db	512 TAANCTATATTTTCANGATG-ACNMAAGATGAAAAATATATNAG-ATTTTTTCTGTGCT	455		
OY	146 TTTTCTCTGTTTCAGATAAAGGAGATGGGTGAGATGCATAGGGAACACTCATATACAT	205		
Db	454 TTTTCCNGTTTTCAGATTAAGGAGAGGAGTGAATGCAAA-GGAACTCAATGCATACAT	396		
OY	206 ATAATTTGACATTTATAGAGAGGGAATAAGCAATGAGACACAATTAACAATGTGTGT	265		
Db	395 ATACCTTTGAAGTTTATAGAGAAAGGAAATAGCAAAATGGCCACAATTAACNATGTGTGT	336		
OY	266 GCGTGGGACGAAGACATCTTTGSAAGTCATGAGTTGTTAGTTTAAATCATATATTTGT	325		
Db	335 GCGTGGGACGAAGACATCTTTGSAAGTCATGAGTTGTTAGTTTAAATCATATATTTGT	276		
OY	326 AATAGTGAACCTGTGACACAAATATTAAGCACCTGGAACCTGGCTTTCACATCTTGAAA	385		
Db	275 AATAGTGAANCTGTGACCAAAATATTAAGCACCTGGAANCTGGCTTTCACATCTTGAAA	216		
OY	386 TTTGACCAACAAGTGTCTATATATATGACAGATCTAATGTAAATCCAGAACTTGGACTCAT	445		
Db	215 TTTGACCAACAAGTGTCTATATATATGACAGATCTAATGTAAATCCAGACTTGGACTCAT	156		
OY	446 CGTTAAATATATTTATGCTATACATTCAAATGTCGCTTAATATGCTTCACAGTAA	505		
Db	155 CGTTAAATATATTTATGCTATACATTCAAATGTCGCTTAATATATGCTTCACAGTAA	96		
OY	506 ATCTAAAAAAGCATTTGTGATTGAAAGCTGCTTTCTATTACTTGTGACTGTGACATA	565		
Db	95 ATCTAAAAAAGCATTTGTGATTGAAAGCTGCTTTCTATTACTTGTGACTGTGACATA	36		
OY	566 CATACTTTTATATGAGCTATGAAATAAACATTTT	600		
Db	35 CATGCTTTTATATGAGCTATGAAATAAACATTTT	1		
RESULT 4				
AAZ77543				
ID	AAZ77543	standard: cDNA: 1545 BP.		
XX	AAZ77543;			
AC				
XX				
DT	10-APR-2000	(first entry)		
XX				
DE	Human ovarian tumor cDNA library derived EST fragment 94.			
XX				
KM	Expressed sequence tag; EST: human; ovarian tumor; anticancer;			
KN	gene therapy; treatment; ss.			
XX				
OS	Homo sapiens.			
XX				
PN	DE19817557-A1.			
XX				
PD	21-OCT-1999.			
XX				
PF	09-APR-1998; 98DE-1017557.			
XX				
PR	09-APR-1998; 98DE-1017557.			
XX				
PA	(META-) METAGEN GES GENOMFORSCHUNG MBH.			
XX				
PI	Rosenthal A, Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl E;			

DR WPI: 1999-591920/51.

XX New nucleic acid sequences expressed in ovarian, and some other, cancer
PT tissues, and derived polypeptides, for treatment of ovarian cancer and
PT identification of therapeutic agents

XX Claim 3: Page 219; 310pp: German.

XX This invention describes novel nucleic acid (cDNA) sequences (A) which
CC have anticancer activity and are highly expressed in ovarian tumor
CC tissue (and some also in testis and breast cancer tissue). The products
CC of the invention can be used for gene therapy. (A) are used (1) for
CC recombinant expression of polypeptides (B) and (11) to isolate complete
CC genes. (B) are used (1) to identify agents suitable for treatment of
CC ovarian cancer; (11) directly for treating this form of cancer
CC (including expression from gene therapy vectors) and (111) for generation
CC of specific antibodies. (A) are identified by assembling ESTs (expressed
CC sequence tags) from a particular tissue type before comparison of
CC expression patterns. This allows a significantly longer fragment of the
CC gene to be revealed, so should reduce the number of failures associated
CC with the fact that ESTs from different libraries may represent different
CC parts of the same unknown gene, distorting the estimated frequency of
CC occurrence in a particular tissue. AA77450-277572 represent the human
CC ovarian tumor cDNA library derived EST fragments described in the method
CC of the invention and encode the protein fragments represented in
XX AA76505-Y76638.

XX Sequence 1545 BP; 460 A; 307 C; 369 G; 409 T; 0 other;

XX Query Match 67.2%; Score 451.4; DB 20; Length 1545;

XX Best Local Similarity 99.8%; Pred. No. 1.6e-69;

XX Matches 452; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

QY 156 TCAGATAAAGGAGATGGTGAGATGATGCGAATCATGCTAATCTATATATTGAA 215
DB 1077 TGAGATAAAGGAGATGGTGAGATGATGCGAATCATGCTAATCTATATATTGAA 1136
QY 216 GATTATAGAAGGAGGAAATAGCAATGCAATATGCTGCTGGGAGG 275
DB 1137 GATTATAGAAGGAGGAAATAGCAATGCAATATGCTGCTGGGAGG 1196
QY 276 AAGACATCTTTGAAGCTCATGAGTTGTTAGTTAAATCATATATTTGTAATAGTAA 335
DB 1197 AAGACATCTTTGAAGCTCATGAGTTGTTAGTTAAATCATATATTTGTAATAGTAA 1256
QY 336 CCTGATCTCAAAATATTAACGACCTTGAACCTGCTTACCAATCTTGAATTTGACCA 395
DB 1257 CCTGATCTCAAAATATTAACGACCTTGAACCTGCTTACCAATCTTGAATTTGACCA 1316
QY 396 AGTGCTTATATATGAGATCTAATGTAAATTCAGAACTGGACATCGTTAAATTT 455
DB 1317 AGTGCTTATATATGAGATCTAATGTAAATTCAGAACTGGACATCGTTAAATTT 1376
QY 456 ATTTATGTGTAAATCAATGCTGATTAATATGCTTCCACAGATAAATCTGAAAA 515
DB 1377 ATTTATGTGTAAATCAATGCTGATTAATATGCTTCCACAGATAAATCTGAAAA 1436
QY 516 CCGATTGTGATGAAAGCTGCTTCTATTTACTGATCTTGTATACATACATCTTTT 575
DB 1437 CCGATTGTGATGAAAGCTGCTTCTATTTACTGATCTTGTATACATACATCTTTT 1496
QY 576 TATGAGCTATGAATTAATCAATTTTAACTGAA 608
DB 1497 TATGAGCTATGAATTAATCAATTTTAACTGAA 1529

```

RESULT 5

XX AAF16221 standard; cDNA; 1597 BP.

XX AAF16221;

XX 13-MAR-2001 (first entry)

XX Human prostate cancer antigen nucleotide sequence SEQ ID NO:656.

XX Human; prostate cancer; prostate cancer antigen; detection; diagnosis;
XX neuroprotective; cytostatic; cardioprotective; immunomodulatory; muscular;
XX vulnary; gastrointestinal; nephrotoxic; antiinfective; gynaecological;
XX antibacterial; gene therapy; neural; immune; reproductive; renal;
XX gastrointestinal; pulmonary; cardiovascular; proliferative disorder;
XX wound; infectious disease; ss.

XX Homo sapiens.

XX MO200055174-A1.

XX 21-SEP-2000.

XX 08-MAR-2000; 2000MO-US05988.

XX 12-MAR-1999; 99US-0124270.

XX (HUMA-) HUMAN GENOME SCI INC.

XX (ROSE/) ROSEN C A.

XX Rosen CA, Ruben SM;

XX WPI: 2000-587513/55.

XX P-PSDB: AAB57018.

XX prostate cancer associated gene sequences, referred to as prostate

XX cancer antigens, useful for treatment, prevention, and diagnosis of

XX disorders such as prostate cancer -

XX Claim 1: Page 1095-1096; 2338pp: English.

XX AAF15566 to AAF16505 encode the human prostate cancer associated
XX proteins, called prostate cancer antigens, given in AAB56363 to AAB57302.
XX The prostate cancer antigens can have neuroprotective, cytostatic,
XX cardioprotective, immunomodulatory, muscular, vulnary, gastrointestinal,
XX nephrotoxic, antiinfective, gynaecological and antibacterial activities,
XX and can be used in gene therapy. The prostate cancer antigen
XX polynucleotides may be used for detection of prostate cancer, chromosome
XX identification, as chromosome markers, and for numerous other diagnostic
XX or research purposes. The prostate cancer antigens may be used to treat
XX disorders such as neural, immune, muscular, reproductive,
XX gastrointestinal, pulmonary, cardiovascular, renal, and proliferative
XX disorders, wounds, and infectious diseases. AAF16506 to AAF16514 to
XX AAB57303 represent sequences used in the exemplification of the present
XX invention.

XX Sequence 1597 BP; 477 A; 323 C; 378 G; 418 T; 1 other;

XX Query Match 67.2%; Score 451.4; DB 21; Length 1597;

XX Best Local Similarity 99.8%; Pred. No. 1.7e-69;

XX Matches 452; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

QY 156 TCAGATAAAGGAGATGGTGAGATGATGCGAATCATGCTAATCTATATATTGAA 215
DB 1088 TCAGATAAAGGAGATGGTGAGATGATGCGAATCATGCTAATCTATATATTGAA 1147
QY 216 GATTATAGAAGGAGGAAATAGCAATGCAATATGCTGCTGGGAGG 275
DB 1148 GATTATAGAAGGAGGAAATAGCAATGCAATATGCTGCTGGGAGG 1207
QY 276 AAGACATCTTTGAAGCTCATGAGTTGTTAGTTAAATCATATATTTGTAATAGTAA 335
DB 1208 AAGACATCTTTGAAGCTCATGAGTTGTTAGTTAAATCATATATTTGTAATAGTAA 1267
QY 336 CCTGATCTCAAAATATTAACGACCTTGAACCTGCTTACCAATCTTGAATTTGACCA 395
DB 1268 CCTGATCTCAAAATATTAACGACCTTGAACCTGCTTACCAATCTTGAATTTGACCA 1327
QY 396 AGTGCTTATATATGAGATCTAATGTAAATTCAGAACTGGACATCGTTAAATTT 455

```

Db 1328 AGGCTCTTATATGACGATCTAATGTAAATCCAGAACTTGACCTCAGCTTTAAATTT 1387
 QY 456 ATTATGATGTAACTCAATATGTGATTAATATGCTTCACAGTAATCTGAATAA 515
 Db 1388 ATTTATGTGTACATTCATAATGTGTGATTAATATGCTTCACAGTAATCTGAATAA 1447
 QY 516 CTGATTTGTGATGAAAGCTGCTTTCTATTACTTGAGTCTTGATACATACCTTTT 575
 Db 1448 CTGATTTGTGATGAAAGCTGCTTTCTATTACTTGAGTCTTGATACATACCTTTT 1507
 QY 576 TATGAGCTATGAAATTAACATTTTAAACTGAA 608
 Db 1508 TATGAGCTATGAAATTAACATTTTAAACTGAA 1540

RESULT 6

AAH33540
 ID AAH33540 standard; cDNA: 1597 BP.

AC AAH33540;

DT 03-SEP-2001 (first entry)

DE Human colon cancer antigen encoding cDNA SEQ ID NO:596.

XX Human; colon cancer; colon cancer antigen; diagnosis; detection;

KW colorectal carcinoma; chromosome 4; ss.

OS Homo sapiens.

XX MO200122920-A2.

XX 05-APR-2001.

PE 28-SEP-2000; 2000MO-US26524.

XX 29-SEP-1999; 99US-0157137.

PR 03-NOV-1999; 99US-0163280.

XX (HUMA-) HUMAN GENOME SCI INC.

PI Ruben SM, Barash SC, Birse CE, Rosen CA;

DR WPI: 2001-235357/24.

DR P-PSDB; AAG74109.

XX Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
 PT useful for preventing, diagnosing and/or treating colorectal cancers -
 XX Claim 1; Page 2645-2646; 9803pp; English.

CC AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
 CC cancer-associated nucleic acid molecules (N) and proteins (P), where
 CC the proteins are collectively known as colon cancer antigens. The colon
 CC cancer antigens have cytostatic activity and can be used in gene
 CC therapy and vaccine production. N and P may be used in the prevention,
 CC diagnosis and treatment of diseases associated with inappropriate P
 CC expression. For example, N and P may be used to treat disorders
 CC associated with decreased expression by rectifying mutations or deletions
 CC in a patient's genome that affect the activity of P by expressing
 CC inactive proteins or to supplement the patients own production of P.
 CC Additionally, N may be used to produce the colon cancer-associated Ps,
 CC by inserting the nucleic acids into a host cell and culturing the cell
 CC to express the proteins. N and P can be used in the prevention, diagnosis
 CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
 CC and AAB77789 represent sequences used in the exemplification of the
 CC present invention.
 CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were
 CC missing at time of publication, meaning no sequences are present for
 CC SEQ ID NO:1027 to 1052, 7921 and 7922.

XX Sequence 1597 BP; 477 A; 323 C; 378 G; 418 T; 1 other;

Query Match 67.2%; Score 451.4; DB 22; Length 1597;
 Best Local Similarity 99.8%; Pred. No. 1.7e-69;
 Matches 452; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 156 TCAGTAAAGAGATGGGTGAGATGATGAGGAACTCATGCTATATATATTGAA 215
 Db 1088 TGAGATTAAGAGAGATGGGTGAGATGATGAGGAACTCATGCTATATATTGAA 1147
 QY 216 GATTATAGAGAGAGGAAATGCAAAATGACACAAATTAATGATGTGGTGGAGC 275
 Db 1148 GATTATAGAGAGGAAATGCAAAATGACACAAATTAATGATGTGGTGGAGC 1207
 QY 276 AAGACATCTTGAAGGTGATGATTTAGTTTACATCATATATTTGTAATAGTAA 335
 Db 1208 AAGACATCTTGAAGGTGATGATTTAGTTTACATCATATATTTGTAATAGTAA 1267
 QY 336 CCTGACTCAAAATTAATGACAGCTGAAACCTGCTTACCAATCTTGAATTTGACCA 395
 Db 1268 CCTGACTCAAAATTAATGACAGCTGAAACCTGCTTACCAATCTTGAATTTGACCA 1327
 QY 396 AGTGTCTTATATATGACATCATATGTAATAATCCAGAACTTGACCTCATGCTTAAAT 455
 Db 1328 AGTGTCTTATATATGACATCATATGTAATAATCCAGAACTTGACCTCATGCTTAAAT 1387
 QY 456 ATTTATGTGTAAATCAATTAATGTGATTAATATGCTTCACAGTAATCTGAATAA 515
 Db 1388 ATTTATGTGTAAATCAATTAATGTGATTAATATGCTTCACAGTAATCTGAATAA 1447
 QY 516 CTGATTTGTGATGAAAGCTGCTTTCTATTACTTGAGTCTTGATACATACCTTTT 575
 Db 1448 CTGATTTGTGATGAAAGCTGCTTTCTATTACTTGAGTCTTGATACATACCTTTT 1507
 QY 576 TATGAGCTATGAAATTAACATTTTAAACTGAA 608
 Db 1508 TATGAGCTATGAAATTAACATTTTAAACTGAA 1540

RESULT 7

ABO54273
 ID ABO54273 standard; cDNA: 1653 BP.

XX ABO54273;

DT 22-AUG-2002 (first entry)

DE Human ovarian antigen HMLHW86 cDNA, SEQ ID NO:153.

XX Human; ovarian antigen; ovary; ovarian; breast; cancer; tumour;
 KW ovarian cancer; breast cancer; tumour; reproductive system disorder;
 KW infertility; pregnancy disorder; anovulation; polycystic ovary syndrome;
 KW PCOS; ovarian cyst; dysmenorrhoea; endocrine disorder; infection;
 KW inflammatory condition; immune disorder; blood disorder;
 KW cardiovascular disorder; respiratory disorder; neurological disorder;
 KW gastrointestinal disorder; urinary system disorder; drug screening;
 KW gene therapy; chromosome mapping; forensic analysis;
 KW antibody preparation; cytostatic; immunomodulatory; neuroprotective;
 KW antiinflammatory; gynaecological; reproductive; chromosome 4q;
 KW gene; ss.

XX Homo sapiens.

XX MO200200677-A1.

XX 03-JAN-2002.

XX 07-JUN-2001; 2001MO-US18569.

XX 07-JUN-2000; 2000US-209467P.

XX (HUMA-) HUMAN GENOME SCI INC.

PI Birse CE, Rosen CA;

CC Note: With the exception of SEQ ID No 1 and 2, the sequence data for this patent did not form part of the printed specification but was supplied by the European Patent Office.

Query Match 65.8%; Score 442.4; DB 24; Length 481;
Best Local Similarity 99.6%; Pred. No. 5.7e-68;
Matches 443; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

SO Sequence 481 BP; 149 A; 93 C; 64 G; 174 T; 1 other:

CC 276 AAGACATCTTGAAGTCATGAGTTGTAGTTTAACATCATATATTTGTATAGTGAAA 335
DB 325 AAGACATCTTGAAGTCATGAGTTGTAGTTTAACATCATATATTTGTATAGTGAAA 266
QY 336 CCTGACTCAAAATATATAGCAGCTTGAAGCTGGCTTTACCAATCTTGAATTTGACCA 395
DB 265 CCTGACTCAAAATATATAGCAGCTTGAAGCTGGCTTTACCAATCTTGAATTTGACCA 206
QY 396 AGTGCTATATATATGACATGATTAATGTAATATCCAGAACTGGACATGCTTAAAT 455
DB 205 AGTGCTATATATATGACATGATTAATGTAATATCCAGAACTGGACATGCTTAAAT 146
QY 456 ATTATGTATACATTCAAATGTGTCATTAATATGCTCCACAGTAATCTGAAAA 515
DB 145 ATTATGTATACATTCAAATGTGTCATTAATATGCTCCACAGTAATCTGAAAA 86
QY 516 CTGATTTGTATGAAGCTGCTTTCTATTTACTTGAAGTCTGTACATACATCTTTT 575
DB 85 CTGATTTGTATGAAGCTGCTTTCTATTTACTTGAAGTCTGTACATACATCTTTT 26
QY 576 TATGAGCTATGAATATAAACAATTT 600
DB 25 TATGAGCTATGAATATAAACAATTT 1

RESULT 9
ABK46008/c
ID ABK46008 standard; cDNA: 481 BP.
XX
AC ABK46008;
XX
DT 05-JUN-2002 (first entry)
XX
DE cDNA encoding colon tumour protein, SEQ ID NO.1559.
XX
KW Human; colon tumour; vaccine; colon cancer; immunogenic;
KW Immunotherapy; gene; ss.
XX
OS Homo sapiens.
XX
PN WO200212328-A2.
XX
PD 14-FEB-2002.
XX
PE 31-JUL-2001; 2001WO-US24218.
XX
PR 03-AUG-2000; 2000US-223283P.
PR 28-MAR-2001; 2001US-279763P.
PR 29-JUN-2001; 2001US-302051P.
XX
PA (CORI-) CORIXA CORP.
XX
XX King GE, Meagher MJ, Xu J, Secrist H;
PI
DR WPI; 2002-241739/29.
XX
PT New colon cancer polypeptides and polynucleotides, useful as vaccines,
PT for diagnosing, preventing, and treating colon cancer, and as markers
PT for the progression of cancer -
XX
PS Claim 1; SEQ ID No 1559; 147pp: English.
XX
XX The invention relates to polynucleotides encoding colon tumour proteins.
CC The polynucleotides and encoded polypeptides are useful in pharmaceutical
CC compositions, such as vaccines, for the diagnosis, prevention, and
CC treatment of colon cancer. Polynucleotide sequences may be used as
CC hybridisation probes or primers, and in the design and preparation of
CC ribozyme molecules for inhibiting expression of tumour polypeptides and
CC proteins in tumour cells. The compositions are useful for stimulating an
CC immune response against cancer, particularly for the immunotherapy of
CC colon cancer, and as markers for the progression of cancer.
CC ABK4450-ABK46237 represent coding sequences of human colon tumour
CC proteins of the invention.

CC Note: With the exception of SEQ ID No 1 and 2, the sequence data for this patent did not form part of the printed specification but was supplied by the European Patent Office.

Query Match 65.8%; Score 442.4; DB 24; Length 481;
Best Local Similarity 99.6%; Pred. No. 5.7e-68;
Matches 443; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

SO Sequence 481 BP; 149 A; 93 C; 64 G; 174 T; 1 other:

CC 276 AAGACATCTTGAAGTCATGAGTTGTAGTTTAACATCATATATTTGTATAGTGAAA 335
DB 325 AAGACATCTTGAAGTCATGAGTTGTAGTTTAACATCATATATTTGTATAGTGAAA 266
QY 336 CCTGACTCAAAATATATAGCAGCTTGAAGCTGGCTTTACCAATCTTGAATTTGACCA 395
DB 265 CCTGACTCAAAATATATAGCAGCTTGAAGCTGGCTTTACCAATCTTGAATTTGACCA 206
QY 396 AGTGCTATATATATGACATGATTAATGTAATATCCAGAACTGGACATGCTTAAAT 455
DB 205 AGTGCTATATATATGACATGATTAATGTAATATCCAGAACTGGACATGCTTAAAT 146
QY 456 ATTATGTATACATTCAAATGTGTCATTAATATGCTCCACAGTAATCTGAAAA 515
DB 145 ATTATGTATACATTCAAATGTGTCATTAATATGCTCCACAGTAATCTGAAAA 86
QY 516 CTGATTTGTATGAAGCTGCTTTCTATTTACTTGAAGTCTGTACATACATCTTTT 575
DB 85 CTGATTTGTATGAAGCTGCTTTCTATTTACTTGAAGTCTGTACATACATCTTTT 26
QY 576 TATGAGCTATGAATATAAACAATTT 600
DB 25 TATGAGCTATGAATATAAACAATTT 1

RESULT 10
AAH83718/c
ID AAH83718 standard; cDNA: 545 BP.
XX
AC AAH83718;
XX
DT 25-SEP-2001 (first entry)
XX
DE Human ovarian tumour associated polynucleotide sequence SEQ ID NO:1342.
XX
KW Human; ovarian tumour; ovarian cancer; diagnosis; gene therapy;
KW Immunogenic; vaccine; ss.
XX
OS Homo sapiens.
XX
PN WO200151513-A2.
XX
PD 19-JUL-2001.
XX
PE 16-JAN-2001; 2001WO-US01575.
XX
PR 14-JAN-2000; 2000US-0176722.
XX
PA (CORI-) CORIXA CORP.
XX
PI Algate PA;
XX
DR WPI; 2001-425866/45.
XX
PT Novel ovarian tumor proteins, and nucleic acids encoding them, used to

PT treat and diagnose cancers, particularly ovarian cancer -
 XX Claim 5; Page 308; 338bp; English.
 XX
 CC AAH82377 to AAH83878 represent human ovarian tumour-associated
 CC polynucleotide sequences which encode ovarian tumour proteins. The
 CC ovarian tumour protein and polynucleotide sequences have cytostatic
 CC activity, and can be used in gene therapy and vaccine production. The
 CC ovarian tumour proteins and polynucleotides can be used to inhibit
 CC the development of cancer, particularly ovarian cancer. They can also
 CC be used to diagnose the onset and progression of cancer.
 XX
 SQ Sequence 545 BP; 163 A; 112 C; 89 G; 177 T; 4 other;
 Query Match 65.8%; Score 442.2; DB 22; Length 545;
 Best Local Similarity 96.0%; Pred. No. 6.2e-68;
 Matches 453; Conservative 0; Mismatches 19; Indels 0; Gaps 0;
 QY 129 GAATTTTCTGCTGCTTTCTCTGTTCAATTAAGAGATGGGTGATGATGAGGA 188
 DB 472 GGAATTCGCCNTTAGCGTGGTCCGCCGAGGTAAGAGATGGGTGATGATGAGGA 413
 QY 189 ACTGAATGCATATATATATATTTGAGATTATGAGAGCGAAATAGCAATGACAC 248
 DB 412 ACTGAATGCATATATATATTTGAGATTATGAGAGCGAAATAGCAATGACAC 353
 QY 249 AAATTAACAAATGCTGCTGCTGGACGACAGACATCTTGAAGTCATGATTTGTTAGTT 308
 DB 352 AAATTAACAAATGCTGCTGCTGGACGACAGACATCTTGAAGTCATGATTTGTTAGTT 293
 QY 309 TAACATATATATTTGTAATAGTGAACCTGCTACTCAAAATATAGCAGCTTGAACCTGG 368
 DB 292 TAACATATATATTTGTAATAGTGAACCTGCTACTCAAAATATAGCAGCTTGAACCTGG 233
 QY 369 CTTTACCAATCTTGAATTTGACACAAAGTGTCTTATATATGAGATCTATATGAAATATC 428
 DB 232 CTTTACCAATCTTGAATTTGACACAAAGTGTCTTATATATGAGATCTATATGAAATATC 173
 QY 429 CAGAACTTGACCTCCATCGTTAAATATTTATGTGTACATTCAAATGTGTGCAATTA 488
 DB 172 CAGAACTTGACCTCCATCGTTAAATATTTATGTGTACATTCAAATGTGTGCAATTA 113
 QY 489 TATGCTTCCACAGTAAATCTGAAAACTGATTTGATTTGAAGCTGCTTTCTATTTA 548
 DB 112 TATGCTTCCACAGTAAATCTGAAAACTGATTTGATTTGAAGCTGCTTTCTATTTA 53
 QY 549 CTGAGCTCTGATCATATACATTTTATGAGCTATGAATAAACATTTT 600
 DB 52 CTGAGCTCTGATCATATACATTTTATGAGCTATGAATAAACATTTT 1
 RESULT 11
 AAS25073/C
 ID AAS25073 standard; cDNA; 439 BP.
 XX
 AC AAS25073;
 XX
 DT 07-NOV-2001 (first entry)
 XX
 DE Human ovarian PCR-subtracted cDNA library clone #1254.
 XX
 XX Immunogenic protein; cancer; ovarian tumour; T-cell stimulation; ss;
 KW gene therapy; cytostatic; T-cell expansion; nucleic acid hybridisation;
 KM primer; probe.
 XX
 OS Homo sapiens.
 XX
 PN WO200157207-A2.
 XX
 PD 09-AUG-2001.
 XX
 PF 05-FEB-2001; 2001WO-US03733.
 XX

PR 04-FEB-2000; 2000US-0180403.
 PR 28-MAR-2000; 2000US-0192745.
 XX
 PA (CORI-) CORIXA CORP.
 XX
 PI Algate PA, Mannion J;
 XX
 DR WPI; 2001-488879/53.
 XX
 XX New polynucleotides encoding ovarian tumour proteins, useful for
 PT treating ovarian cancer, and as probes, primers, and markers of cancer
 PT progression -
 XX
 PS Example 1; page 305; 378bp; English.
 XX
 CC The invention comprises compositions used for the therapy and diagnosis
 CC of ovarian cancer. The compositions comprise one or more ovarian tumour
 CC proteins, their associated polynucleotides, or immunogenic portions of
 CC the proteins. The ovarian tumour polynucleotides and polypeptides are
 CC useful for stimulating and/or expanding T cells specific for a tumour
 CC protein. They are also useful for inhibiting the development of cancer in
 CC a patient with an ovarian tumour DNA or protein by incubating isolated
 CC T-cells allowing them to proliferate, and administering to the patient.
 CC The sequences can be used as markers for cancer, for example, to monitor
 CC ovarian cancer progression. Probes and primers are useful in nucleic acid
 CC hybridisation, in detecting the presence of complementary sequences in a
 CC given sample, for preparing mutant species and for preparing other
 CC genetic constructions. Sequences AAS22820-AAS25231 and AAS25238-AAS25549
 CC represent human ovarian tumour protein cDNA clones.
 XX
 SQ Sequence 439 BP; 142 A; 80 C; 60 G; 157 T; 0 other;
 Query Match 65.3%; Score 439; DB 22; Length 439;
 Best Local Similarity 100.0%; Pred. No. 2.2e-67;
 Matches 439; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 162 AAGAGAGATGGGTGAGATGATAGGAACTCAATGCAATATATTAATTTGAAGATTAT 221
 DB 439 AAGAGAGATGGGTGAGATGATAGGAACTCAATGCAATATATTAATTTGAAGATTAT 380
 QY 222 AGAAGAGGGAATAGCAATGACACAAATTTCAAAATGCTGTGGTGGAGGAGACA 281
 DB 379 AGAAGAGGGAATAGCAATGACACAAATTTCAAAATGCTGTGGTGGAGGAGACA 320
 QY 282 TCTTTGAAGTCAATGATTTGTTAGTTTACATCATATATTTGTAATAGGAAACCTGTA 341
 DB 319 TCTTTGAAGTCAATGATTTGTTAGTTTACATCATATATTTGTAATAGGAAACCTGTA 260
 QY 342 CTCAAAATATATAGCAGCTTGAACCTGCTTACCAATCTTGAATTTGACACAGTGTG 401
 DB 259 CTCAAAATATATAGCAGCTTGAACCTGCTTACCAATCTTGAATTTGACACAGTGTG 200
 QY 402 TTTATATATGAGATCTATATTTAAATTCACAGACTTGAGCTCCATCGTTAAATATTTAT 461
 DB 199 TTTATATATGAGATCTATATTTAAATTCACAGACTTGAGCTCCATCGTTAAATATTTAT 140
 QY 462 GTGTAACATTCAAATGCTGATTAATATATGCTTCCACAGATAAATCTGAAAAAGTATT 521
 DB 139 GTGTAACATTCAAATGCTGATTAATATATGCTTCCACAGATAAATCTGAAAAAGTATT 80
 QY 522 TGTGATTTGAAGCTGCTTCTATTTACTTGAAGCTTGTATCATATACTTTTATGAG 581
 DB 79 TGTGATTTGAAGCTGCTTCTATTTACTTGAAGCTTGTATCATATACTTTTATGAG 20
 QY 582 CTATGAATTAACATTTT 600
 DB 19 CTATGAATTAACATTTT 1
 RESULT 12
 AAT29042/C
 ID AAT29042 standard; cDNA; 425 BP.
 XX

AC AAI29042;
XX
DT 12-OCT-2001 (first entry)
XX
XX Colon tumour related determined cDNA sequence for clone CS1-160.
DE
XX
XX Human; immunotherapy; diagnosis; colon cancer; colon tumour;
KM immunogenic; gene therapy; vaccine; colonic cancer; ss.
XX
OS Homo sapiens.
PN MO200149716-A2.
PD 12-JUL-2001.
XX
PF 29-DEC-2000; 2000WO-US35596.
XX
PR 30-DEC-1999; 99US-0476296.
PR 10-JAN-2000; 2000US-0480321.
PR 15-FEB-2000; 2000US-0504629.
PR 06-MAR-2000; 2000US-0519444.
PR 19-MAY-2000; 2000US-0575251.
PR 29-JUN-2000; 2000US-0609448.
PR 28-AUG-2000; 2000US-0649811.
XX
PA (CORI-) CORIXA CORP.
PI Xu J, Lodes MJ, Secret H, Benson DR, Weagher MJ, Stolk JA;
PI King GE, Wang T, Jlang Y;
XX
XX WPI: 2001-441847/47.
PT Colon tumor associated proteins and nucleic acids useful for the
PT prevention, diagnosis and treatment of colonic cancer -
XX
XX
PS Claim 25; Page 291; 472pp; English.
XX
XX The present invention describes colon tumour associated proteins (I) and
XX the polynucleotides (II) that encode them. (I) have cytostatic activity.
XX (I) and (II) can be used in gene therapy and vaccine production. (I) and
XX (II) may be used in the prevention, diagnosis and treatment of diseases
XX associated with inappropriate colon tumour associated protein (TCAP)
XX expression, such as colonic cancer. For example, (I) and (II) may be
XX used to treat disorders associated with decreased expression by
XX rectifying mutations or deletions in a patient's genome that affect the
XX activity of TCAPs by expressing inactive proteins or to supplement the
XX patients own production of them. Additionally, (II) may be used to
XX produce the TCAP proteins, by inserting the nucleic acids into a host
XX cell culturing the cell to express the protein. (II) and its
XX complementary sequences may also be used as DNA probes in diagnostic
XX polymerase chain reaction (PCR) and hybridisation assays to detect and
XX quantitate the presence of similar nucleic acids in samples, and
XX therefore which patients may be in need of restorative therapy. (I) may
XX also be used as antigens in the production of antibodies against TCAPs
XX and in assays to identify modulators of TCAP expression and activity.
XX Anti-(I) antibodies and antagonists may also be used to down regulate
XX TCAP expression and activity. The anti-(I) antibodies may also be used
XX as diagnostic agents for detecting the presence of TCAPs in samples
XX (e.g. by enzyme linked immunosorbant assay (ELISA)). AAI28460 to AAI29512
XX and AAI24494 to AAI24523 represent nucleotide and amino acid sequences
XX given in the exemplification of the present invention.
XX
SQ Sequence 425 BP; 132 A; 83 C; 59 G; 151 T; 0 other;

Query Match 61.88; Score 415.4; DB 22; Length 425;
Best Local Similarity 99.88; Pred. No. 2.7e-63;
Matches 416; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 156 TCAGATAAGGAGATGGCTGAGATGCATAGGAACTCATGATATATATTTGAA 215
D 417 TGAGATTAAGGAGATGGCTGAGATGCATAGGAACTCATGATATATATTTGAA 358
OY 216 GATTATAGAGAGGAAATAGCAAAATGACACAAATTAACAATGTGTGCTGGGACG 275

DB 357 GATTATAGAGAGGAAATAGCAAAATGACACAAATTAACAATGTGTGCTGGGACG 298
QY 276 AAGACATCTTTGAAGGTCATGAGTTGTAGTTAAATCATATATTTGTAATAGTAAA 335
DB 297 AAGACATCTTTGAAGGTCATGAGTTGTAGTTAAATCATATATTTGTAATAGTAAA 238
QY 336 CCTGACTCAAAATATATAGCAGCTTGAACGCTTACCAATCTTGAATTTGACCA 395
DB 237 CCTGACTCAAAATATATAGCAGCTTGAACGCTTACCAATCTTGAATTTGACCA 178
QY 396 AGTGCTTATATATGACATCTAATGTAAATCCGAACTGAGATCCATGTTAAATTT 455
DB 177 AGTGCTTATATATGACATCTAATGTAAATCCGAACTGAGATCCATGTTAAATTT 118
QY 456 ATTTATGTGTAAACATTCATATGTGTGATTAATATGCTTCCACAGTAAATCTGAAAA 515
DB 117 ATTTATGTGTAAACATTCATATGTGTGATTAATATGCTTCCACAGTAAATCTGAAAA 58
QY 516 CTGATTTGTGATGAAACGCTTCTTATTTACTTGTGATCTTGTACATACATCT 572
DB 57 CTGATTTGTGATGAAACGCTTCTTATTTACTTGTGATCTTGTACATACATCT 1

RESULT 13
ABK36208
ID ABK36208 standard; CDNA: 1439 BP.
XX
XX ABK36208;
AC
XX
XX 08-MAY-2002 (first entry)
DT
XX
XX CDNA sequence #599 encoding novel human secreted protein.
DE
XX
XX Human secreted protein; hyperproliferative disorder; autoimmune disorder;
XX immune deficiency disorder; blood disorder; inflammatory disorder;
XX infectious disorder; allergic condition; neurodegenerative disorder;
XX liver fibrosis; coagulation disorder; gene therapy; antileprobal;
XX tumour; cancer; hepatotropic; immunosuppressive; antirheumatic; gene; ss.
XX
XX Homo sapiens.
OS
XX
XX WO200177289-A2.
PN
XX
XX 18-OCT-2001.
PD
XX
XX 29-MAR-2001; 2001WO-US10232.
PF
XX
XX 06-APR-2000; 2000US-195605P.
PR
XX
XX (GENY) GENETICS INST INC.
PA
XX
XX Jacobs K, McCoy JM, Lavalie ER, Collins-Racie LA, Evans C;
PI Medberg D, Treacy M, Agostino MJ, Bowman MR, Spaulding V, Wong GG;
PI Clark HF, Fechtel K, Howes SH, Resnick RJ, Guzikola K, Graham JR;
XX
XX WPI: 2002-179322/23.
DR
XX
XX Six hundred and twenty three polynucleotides derived from a variety of
XX human tissue sources which encode secreted proteins, useful for
XX treating immune deficiencies and disorders such as autoimmune disorders
XX
XX
PS Claim 1; Page 384; 393pp; English.
XX
XX The present invention relates to the isolation of novel cDNA sequences
XX which encode human secreted proteins. The cDNA sequences have been
XX derived from a variety of human tissues. The invention also provides
XX a method for producing proteins from these polynucleotide sequences.
XX The proteins are useful for identifying compounds that modulate their
XX activity and production. The sequences of the invention are
XX useful for treating diseases such as hyperproliferative disorders
XX (e.g. cancer), immune deficiency disorders (e.g. severe combined

CC immunodeficiency (SCID), autoimmune disorders (e.g. multiple
 CC sclerosis), blood disorders (e.g. thrombocytopenia), inflammatory
 CC disorders (e.g. arthritis), infectious disorders (e.g. hepatitis),
 CC allergic conditions (e.g. asthma), neurodegenerative disorders (e.g.
 CC Alzheimer's disease), liver fibrosis, coagulation disorders (e.g.
 CC haemophilia), and tumours. The polynucleotide sequences of the
 CC invention are also useful in gene therapy. ABK3610-ABK3623 represent
 CC the cDNA sequences of the invention that encode for novel human
 CC secreted proteins.
 CC
 XX
 SQ Sequence 1439 BP; 417 A; 301 C; 351 G; 370 T; 0 other;

Query Match 53.8%; Score 361.4; DB 24; Length 1439;
 Best Local Similarity 99.7%; Pred. No. 5,9e-54;
 Matches 362; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 156 TCAGATAAAGAGATGGTGGATGATGATAGGAGACATCAGTACTATATTAATTGAA 215
 DB 1077 TGAAGATAAAGAGATGGTGGATGATGATAGGAGACATCAGTACTATATTAATTGAA 1136
 OY 216 GATTATAGAGAAGGAGAAATAGCAATGACACAAATGCTGTGGGAGC 275
 DB 1137 GATTATAGAGAAGGAGAAATAGCAATGACACAAATGCTGTGGGAGC 1196
 OY 276 AAGACATCTTTGAAGGTCATGAGTTGTTAGTTAAATCATATATTTGTAATAGTGAAA 335
 DB 1197 AAGACATCTTTGAAGGTCATGAGTTGTTAGTTAAATCATATATTTGTAATAGTGAAA 1256
 OY 336 CCTGTACTCAAAATATAGCAGCTTGAACCTGGCTTACCAATCTTGAATTTGACCACA 395
 DB 1257 CCTGTACTCAAAATATAGCAGCTTGAACCTGGCTTACCAATCTTGAATTTGACCACA 1316
 OY 396 AGTGTCTTATATATGAGATCTATATGTAATAATCCAGACTGGACATCGTTAAATT 455
 DB 1317 AGTGTCTTATATATGAGATCTATATGTAATAATCCAGACTGGACATCGTTAAATT 1376
 OY 456 ATTATATGCTAACATTCAAATGTGTGCAATTAATATGCTTCCACAGATAAATCTGAAAA 515
 DB 1377 ATTATATGCTAACATTCAAATGTGTGCAATTAATATGCTTCCACAGATAAATCTGAAAA 1436
 OY 516 CTG 518
 DB 1437 CTG 1439

RESULT 14
 AAC69778
 ID AAC69778 standard; cDNA; 536 BP.
 XX
 AC AAC69778;
 XX
 DT 30-JAN-2001 (first entry)
 XX
 DE Human breast tumour antigen partial coding sequence #77.
 XX
 KW Cytostatic; vaccine; human; breast tumour; antigen; breast cancer; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200052165-A2.
 XX
 PD 08-SEP-2000.
 XX
 PF 29-FEB-2000; 2000MO-US05431.
 XX
 PR 04-MAR-1999; 99US-0262505.
 PR 19-MAR-1999; 99US-0272886.
 PR 17-SEP-1999; 99US-0396313.
 XX
 PA (CORI-) CORIXA CORP.
 XX
 PI Lodes MJ;
 XX

DR WP1: 2000-572184/53.
 XX Breast tumor antigen polypeptides and polynucleotides, useful for
 PT manufacturing vaccines and compositions for treating, diagnosing, and
 PT monitoring breast cancer.
 XX
 PS Claim 16; Page 128; 140pp; English.

The present invention relates to immunogenic portions of new human
 CC breast tumor antigens (AAB28183-B28214) and their coding sequences
 CC (AAC69645-C69804). The breast tumor antigen polypeptides of the present
 CC invention and their coding sequences are useful for inhibiting the
 CC development of breast cancer in a patient. The breast tumor antigen
 CC polypeptides and polynucleotides may be used in vaccines and
 CC pharmaceutical compositions for treating breast cancer, and for
 CC diagnosing and monitoring the cancer. The present sequence is a coding
 CC sequence for the immunogenic portion for one such human breast cancer
 CC tumour antigen.
 CC
 XX

SQ Sequence 536 BP; 172 A; 78 C; 127 G; 159 T; 0 other;

Query Match 51.4%; Score 345.4; DB 21; Length 536;
 Best Local Similarity 99.7%; Pred. No. 3.3e-51;
 Matches 346; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 156 TCAGATAAAGAGATGGTGGATGATGATAGGAGACATCAGTACTATATTAATTGAA 215
 DB 113 TGAAGATAAAGAGATGGTGGATGATGATAGGAGACATCAGTACTATATTAATTGAA 172
 OY 216 GATTATAGAGAAGGAGAAATAGCAATGACACAAATGCTGTGGGAGC 275
 DB 173 GATTATAGAGAAGGAGAAATAGCAATGACACAAATGCTGTGGGAGC 232
 OY 276 AAGACATCTTTGAAGGTCATGAGTTGTTAGTTAAATCATATATTTGTAATAGTGAAA 335
 DB 233 AAGACATCTTTGAAGGTCATGAGTTGTTAGTTAAATCATATATTTGTAATAGTGAAA 292
 OY 336 CCTGTACTCAAAATATAGCAGCTTGAACCTGGCTTACCAATCTTGAATTTGACCACA 395
 DB 293 CCTGTACTCAAAATATAGCAGCTTGAACCTGGCTTACCAATCTTGAATTTGACCACA 352
 OY 396 AGTGTCTTATATATGAGATCTATATGTAATAATCCAGACTGGACATCGTTAAATT 455
 DB 353 AGTGTCTTATATATGAGATCTATATGTAATAATCCAGACTGGACATCGTTAAATT 412
 OY 456 ATTATATGCTAACATTCAAATGTGTGCAATTAATATGCTTCCACAGT 502
 DB 413 ATTATATGCTAACATTCAAATGTGTGCAATTAATATGCTTCCACAGT 459

RESULT 15
 AAS25360
 ID AAS25360 standard; cDNA; 507 BP.
 XX
 AC AAS25360;
 XX
 DT 07-NOV-2001 (first entry)
 XX
 DE Human ovarian PCR-subtracted cDNA library clone #1445.
 XX
 KW Immunogenic protein; cancer; ovarian tumour; T-cell stimulation; ss;
 KW gene therapy; cytosstatic; T-cell expansion; nucleic acid hybridisation;
 KW primer; probe.
 XX
 OS Homo sapiens.
 XX
 PN WO200157207-A2.
 XX
 PD 09-AUG-2001.
 XX
 PF 05-FEB-2001; 2001MO-US03733.
 PR 04-FEB-2000; 2000US-0180403.
 XX

PR 28-MAR-2000; 2000US-0192745.

XX (CORI-) CORIXA CORP.

XX Algate PA, Mannion J;

XX WPI; 2001-488879/53.

XX
PT New polynucleotides encoding ovarian tumour proteins, useful for
PT treating ovarian cancer, and as probes, primers, and markers of cancer
PT progression

XX
PS Example 1; page 349; 378pp; English.

CC The invention comprises compositions used for the therapy and diagnosis
CC of ovarian cancer. The compositions comprise one or more ovarian tumour
CC proteins, their associated polynucleotides, or immunogenic portions of
CC the proteins. The ovarian tumour polynucleotides and polypeptides are
CC useful for stimulating and/or expanding T cells specific for a tumour
CC protein. They are also useful for inhibiting the development of cancer in
CC a patient with an ovarian tumour DNA or protein by incubating isolated
CC T-cells allowing them to proliferate, and administering to the patient.
CC The sequences can be used as markers for cancer, for example, to monitor
CC ovarian cancer progression. Probes and primers are useful in nucleic acid
CC hybridisation, in detecting the presence of complementary sequences in a
CC given sample, for preparing mutant species and for preparing other
CC genetic constructions. Sequences AAS23820-AAS25231 and AAS25328-AAS25549
CC represent human ovarian tumour protein CDNA clones.

XX
SQ Sequence 507 BP; 174 A; 64 C; 105 G; 143 T; 21 other;

Query Match 50.0%; Score 335.8; DB 22; Length 507;

Best Local Similarity 98.9%; Pred. No. 1.5e-49;

Matches 348; Conservative 0; Mismatches 3; Indels 1; Gaps 1;

OY 156 TCAGATAAGAGATGGTGCATGAGTCAATGCACTCAATGATATATATTTGAA 215
DB 100 TGAGATTAAGAGATGGTGCATGAGTCAATGCACTCAATGATATATTTGAA 159
OY 216 GATTATAGAGAGGGAATAGCAATGACACAATTAATGATGCTGGTGAGC 275
DB 160 GATTATAGAGAGGGAATAGCAATGACACAATTAATGATGCTGGTGAGC 219
OY 276 AAGACATCTTGAAGTCAATGAGTGTGTTAATCATATATTTGATAGTGA 335
DB 220 AAGACATCTTGAAGTCAATGAGTGTGTTAATCATATATTTGATAGTGA 279
OY 336 CCTGACTCAAAATATTAAGACCTTGAACCTGCTTTACCAATCTTGAAATTTGACCACA 395
DB 280 CCTGACTCAAAATATTAAGACCTTGAACCTGCTTTACCAATCTTGAAATTTGACCACA 339
OY 396 AGTGCTATATATGATGATTAATGATAAATCCAGAACTTGACTCCAT-CGTTAAAT 454
DB 340 AGTGCTATATATGATGATTAATGATAAATCCAGAACTTGACTCCATCGTTAAAT 399
OY 455 TATTTATGTAACATTCGAATGTGTCATTAATATGCTTCACAGTAAA 506
DB 400 TATTTATGTAACATTCGAATGTGTCATTAATATGCTTCACAGTAAA 451

Search completed: November 26, 2002, 20:12:07
Job time : 111.219 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 26, 2002, 19:45:25 : Search time 629.416 Seconds
(without alignments)
17291.221 Million cell updates/sec

Title: US-09-919-497-32

Perfect score: 672
Sequence: 1 aatactattcttctttt.....tttttttttttttggaga 672

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

EST: *
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hnc:*
9: gb_estl:*
10: gb_est2:*
11: gb_hnc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pla:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rdg:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	503.8	75.0	554	9	AI027882
C 2	452.8	67.4	471	10	AM769100
C 3	451.4	67.2	501	12	BG292290
C 4	451.4	67.2	573	13	BI438810
C 5	451.4	67.2	592	13	BM272151
C 6	451.4	67.2	604	10	AM473340

C 7	451.4	67.2	614	9	AI624711
C 8	451.4	67.2	627	9	AI955311
C 9	451.4	67.2	675	9	AI978632
C 10	451.4	67.2	709	14	BM981479
C 11	451.4	67.2	768	9	AI571882
C 12	450.4	67.0	513	9	AA502398
C 13	449.8	66.9	471	9	AI038566
C 14	449.8	66.9	481	13	BM510869
C 15	449.8	66.9	535	10	BE672486
C 16	449.8	66.9	601	10	AM513282
C 17	449.8	66.9	683	9	AI962963
C 18	449.8	66.9	690	10	AM188163
C 19	449.4	66.9	511	9	AA424571
C 20	449.4	66.9	561	9	AI055306
C 21	449.4	66.9	570	10	AM780425
C 22	449.4	66.9	684	10	AM004850
C 23	449.4	66.9	686	9	AI628538
C 24	449.4	66.9	764	10	AM103479
C 25	449.4	66.9	830	9	AI718512
C 26	449.2	66.8	506	9	AI285353
C 27	448.8	66.8	520	10	AM473388
C 28	448.8	66.8	618	10	AM272287
C 29	448.8	66.8	694	10	AM129970
C 30	448.4	66.7	542	9	AI241214
C 31	448.4	66.7	598	14	BQ416873
C 32	448.4	66.7	603	14	BQ778413
C 33	448.4	66.7	612	14	BQ778448
C 34	448.2	66.7	551	9	AI024785
C 35	447.8	66.6	707	10	AM007667
C 36	447.4	66.6	516	9	AI285185
C 37	447.2	66.5	681	9	AI697298
C 38	447.2	66.5	752	9	AI745085
C 39	443.4	66.0	692	9	AI712127
C 40	441.8	65.7	468	9	AA443854
C 41	440.2	65.5	605	10	AM873982
C 42	439.8	65.4	453	9	AI074078
C 43	437.8	65.1	579	9	AA130871
C 44	437.4	65.1	486	9	AI824537
C 45	437.2	65.1	518	10	AM075441

ALIGNMENTS

RESULT 1
AI027882/c
LOCUS
DEFINITION
AI027882
554 bp. mRNA linear EST 28-AUG-1998
IMAGE:1650462.3' similar to gb:U14758 MAJOR GASTROINTESTINAL
TUMOR-ASSOCIATED PROTEIN GA733-2 (HUMAN);, mRNA sequence.

ACCESSION
AI027882
VERSION
AI027882.1 GI:3246581

KEYWORDS
SOURCE
ORGANISM

Human sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE
1 (bases 1 to 554)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgaps-remail.nih.gov
CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo
, Ph.D.

CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bdip/image/image.html
Insert Length: 1292 Std Error: 0.00
Seq primer: -40ml3 fwd. EY from Amersham

FEATURES High quality sequence stop: 455.
 source Location/Qualifiers
 1..554
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:1650462"
 /clone_lib="Soares.parathyroid.tumor_NBHPA"
 /tissue_type="parathyroid tumor"
 /dev_stage="adult"
 /lab_host="DH10B (ampicillin resistant)"
 /note="Organ: parathyroid gland; Vector: pT73D (Pharmacia) with a modified polylinker; site_1: Not I; site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer
 [5'-TGTTACCAATCGAAGTGGAGCGGCCGCCACCAATTTTTTTTTTTTTTTT
 TTTT-3'] , double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT73 vector (Pharmacia). Library went through one round of normalization to a Cot = 5. Library constructed by Bento Soares and M. Fatima Bonaldo. RNA from sporadic parathyroid adenomas was kindly provided by Dr. Stephen Marx, National Institute of Diabetes and Digestive and Kidney Diseases, NIH."

BASE COUNT 196 a 89 c 71 g 198 t
 ORIGIN

Query Match 75.0%; Score 503.8; DB 9; Length 554;
 Best Local Similarity 99.4%; Pred. No. 9e-64;
 Matches 516; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

QY 1 AATATACCTTTCTTTTATTTAAATAGCTTATTAATGGGAAAAATATCTT 60
 |||||
 DB 523 AATATACCTTTCTTTTATTTAAATAGCTTATTAATGGGAAAAATATCTT 464
 QY 61 GTGTCTCTTAATTCATTTTATTAATAGCTATTTGAGAAATGACAAAAGATTGAAA 120
 |||||
 DB 463 GTGTCTCTTAATTCATTTTATTAATAGCTATTTGAGAAATGACAAAAGATTGAAA 404
 QY 121 AATATAGCAATTTTTCGTGCTTTTCCTGTTTCAGATAAAGAGATGGGTGAGATG 180
 |||||
 DB 403 AATATAGCAATTTTTCGTGCTTTTCCTGTTTCAGATAAAGAGATGGGTGAGATG 345
 QY 181 CATAGGAACCTCAATGATATATATTAATTTAGAGATTATAGAGAGGAATATAGCA 240
 |||||
 DB 344 CATAGGAACCTCAATGATATATATTAATTTAGAGATTATAGAGAGGAATATAGCA 265
 QY 241 ATGAGACAAATTAACAATGTGTGCGTGGAAGACATCTTTGAAAGGTGATGAGTT 300
 |||||
 DB 284 ATGAGACAAATTAACAATGTGTGCGTGGAAGACATCTTTGAAAGGTGATGAGTT 225
 QY 301 TGTATTGTTAATCATATATTTGTATATAGTGAACCTGTACCAAAATTAAGCAGCTT 360
 |||||
 DB 224 TGTATTGTTAATCATATATTTGTATATAGTGAACCTGTACCAAAATTAAGCAGCTT 165
 QY 361 GAAAGTGGCTTTTGCAATCTTGAAATTTGACCAAGTGTCTATATATATCAGATTAAT 420
 |||||
 DB 164 GAAAGTGGCTTTTGCAATCTTGAAATTTGACCAAGTGTCTATATATATCAGATTAAT 105
 QY 421 GTAATATCCAGAACTTGAGCTCATGTTAAATTTATTTATGTGTAACTTCAATGTGT 480
 |||||
 DB 104 GTAATATCCAGAACTTGAGCTCATGTTAAATTTATTTATGTGTAACTTCAATGTGT 45
 QY 481 GCATTAAATATGCTTCACAGTAATAATCTGAAAAACTGA 519
 |||||
 DB 44 GCATTAAATATGCTTCACACACAAATCTGAAAAACTGA 6

RESULT 2
 AW769100/c 471 bp mRNA linear EST 04-MAY-2000
 LOCUS h154a02.x1 NCI-CGAP-co14 Homo sapiens cDNA clone IMAGE:3004970 3'
 DEFINITION similar to gb:x14758 MAJOR GASTROINTESTINAL TUMOR-ASSOCIATED

PROTEIN GA733-2 (HUMAN); mRNA sequence.
 ACCESSION AW769100
 VERSION AW769100.1 GI:7701127
 KEYWORDS EST.
 SOURCE
 ORGANISM Homo sapiens
 human.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 1 (bases 1 to 471)
 NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncigap>
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
 Unpublished (1997)
 JOURNAL
 COMMENT
 Contact: Robert Strausberg, Ph.D.
 Email: cgapsb@remail.nih.gov
 Tissue Procurement: Christopher Moskalko, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL, send email to: info@image.llnl.gov
 Seq primer: -40UP from Gibco
 High quality sequence stop: 395.
 Location/Qualifiers
 1..471
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:3004970"
 /clone_lib="NCI-CGAP-co14"
 /tissue_type="moderately-differentiated adenocarcinoma"
 /lab_host="DH10B"
 /note="Organ: colon; Vector: pCMV-SPORT6; Site_1: SalI; Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.7 Kb. Life Technologies catalog #: 11531-019"

BASE COUNT 148 a 87 c 62 g 174 t
 ORIGIN

Query Match 67.4%; Score 452.8; DB 10; Length 471;
 Best Local Similarity 99.6%; Pred. No. 2.e-56;
 Matches 454; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 153 GTTTCAGATTAAGAGATGGGTGATGCAATGAGGACCTCAATGCAATATATTT 212
 |||||
 DB 459 GCTTGAGATTAAGAGATGGGTGATGCAATGAGGACCTCAATGCAATATATTT 400
 QY 213 GAAGATTATAGAAAGAGAAATAGCAAAATGACACAAATTAACAATGTGTGCGTGG 272
 |||||
 DB 399 GAAGATTATAGAAAGAGAAATAGCAAAATGACACAAATTAACAATGTGTGCGTGG 340
 QY 273 ACGAAGACATCTTGGAAGTCAATGAGTGTGTACTTAAATCATATATTTGTAATGTG 332
 |||||
 DB 339 ACGAAGACATCTTGGAAGTCAATGAGTGTGTACTTAAATCATATATTTGTAATGTG 280
 QY 333 AAACCTGACCAAAATATATAGAGCTTGAACCTGAGCTTACCAATTTGGAATTTGAC 392
 |||||
 DB 279 AAACCTGACCAAAATATATAGAGCTTGAACCTGAGCTTACCAATTTGGAATTTGAC 220
 QY 393 ACAAGTCTTATATATGAGATCTAATGTAAATCCAGAACTTGCACTCATCGTTAA 452
 |||||
 DB 219 ACAAGTCTTATATATGAGATCTAATGTAAATCCAGAACTTGCACTCATCGTTAA 160
 QY 453 ATATTATATGTTGATCAATTAATGTGTGATTAATATGCTTCCACAGTAAATCTGAA 512
 |||||
 DB 159 ATATTATATGTTGATCAATTAATGTGTGATTAATATGCTTCCACAGTAAATCTGAA 100
 QY 513 AAATGATTTGATTAAGAGTGCCTTCTATTTATTTGAGCTGTACATACATCTT 572
 |||||
 DB 99 AAATGATTTGATTAAGAGTGCCTTCTATTTATTTGAGCTGTACATACATCTT 40
 QY 573 TTTTATGAGCTATGAATAAACAATTTTAAACTGAA 608

Dn 39 TTTTATGAGCTATGAAATAAACAATTTTAAACTGAA 4
|||||
|||

RESULT 3
LOCUS BG292290
DEFINITION BG292290 Homo sapiens CDNA clone IMAGE:4515466 5', mRNA sequence.
ACCESSION BG292290
VERSION BG292290.1 GI:13050959
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 501)
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.

Plate: LLAM10405 row: f column: 11
High quality sequence stop: 501.

FEATURES	Location/Qualifiers
source	1. .501

```

/organism="Homo sapiens"
/bb_xref="taxon:9606"
/clone_image="IMAGE:4515466"
/clone_lib="NIH_MGC_93"
/tissue_type="transitional cell papilloma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: Bladder; Vector: pCMV-Sport6; Site_1: Notti; Site_2: Sall; Cloned unidirectionally; oligo-dt primed. Average insert size 1.7 Kb. Library enriched for full-length clones and constructed by Life Technologies. Note: this is a NIH_MGC Library."

```

BASE COUNT	195 a	63 c	95 g	148 t	
ORIGIN					
Query Match		67.2%	Score 451.4;	DB 12;	Length 501;
Best Local Similarity		99.8%	Pred. No. 3.3e-56;		
Matches 452;	Conservative	0;	Mismatches	1;	Indels 0;
					Gaps 0;

Qy	156	TCACATTAAGAGAGATGGGTGAGATGCATATAGGAACTCAATCCATATATTAATTTGAA	215
Db	33	TGACATTAAGAGAGATGGGTGAGATGCATATAGGAACTCAATCCATATATTAATTTGAA	92
Qy	216	GATTATAGAGAAAGGGAATATGCATAATGACCAAATTTACAATGTGTGCTGGGACG	277
Db	93	GATTATAGAGAAAGGGAATATGCATAATGACCAAATTTACAATGTGTGCTGGGACG	152
Qy	276	AAGACATCTTTGAAGGTCATGAGTTGTTAGTTAAACATCATATATTTGTAATGTGAAA	335
Db	153	AAGACATCTTTGAAGGTCATGAGTTGTTAGTTAAACATCATATATTTGTAATGTGAAA	212
Qy	336	CCTGTACTCAAAATATATTAAGCAGCTTGAACCTGCTTTACCAATCTTGAAATTTGACCACA	399
Db	213	CCCTGTACTCAAAATATATTAAGCAGCTTGAACCTGCTTTACCAATCTTGAAATTTGACCACA	272
Qy	396	AGTGTCTTAATATATGCAAGATCTAATGTAAAAATCCAGAACTTGAGCTCCATCGTAAAAATT	455
Db	273	AGTGTCTTAATATATGCAAGATCTAATGTAAAAATCCAGAACTTGAGCTCCATCGTAAAAATT	333
Qy	456	ATTATATGTGTAACATTCAAAATGTGTGCATTAAATATATGCTTCCACAGTAAAAATCTGAAAAA	515

Db	333	ATTATGTCGTAACATTCAAAAGTGTGCATTTAAATATGCTTCCACAGTAATAATGTGAAAA	392
Qy	516	CTGATTTGTGATTTGAAGGCGCCCTTTCATTTACTTASGCTGTACATACATACCTTTT	575
Dd	393	CTGATTTGTGATTTGAAGGCGCCCTTTCATTTACTTASGCTGTACATACATACCTTTT	452
Qy	576	TATGAGTCATGAATTAACAATTTTAACGCGAA	608
Dd	453	TATGAGTCATGAATTAACAATTTTAACGCGAA	485

RESULT 4	BI438810/c	LOCUS	DEFINITION
	BI438810	573 bp	mrna
	IC77a04.x3	HR85 1stlet	Homo sapiens CDNA 3, similar to SW-63732_HUMAN
	P16422	MAJOR GASTROINTESTINAL TUMOR-ASSOCIATED PROTEIN	GA733-2
	PRECUSOR ;	mrna sequence.	

ACCESSION	BI438810
VERSION	BI438810.1
REVISION	GI:15263500
EST	

KEYWORDS
SOURCE

ORGANISM

REFERENCE	AUTHORS
Mammalia; Eutheria; 1 (bases 1 to 573)	Primates; Catarrhini; Hominoidea; Homo.
Melton, D., Brown, J., Kenty, G., Permutt, A., Lee, C., Kaestner, K.,	

TITLE

JOURNAL
COMMENT

Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge
MA 02138

Tel: 617-495-1812
Fax: 617-495-8557

Email: dmelton@biohp.harvard.edu
Library was constructed by Dr. Hiroshi Inoue DNA sequencing by:

Washington University Genome Sequencing Center For Information on obtaining a clone please contact: Dr. Hiroshi Inoue

containing a stone piece connect. Dr. HINOUKI HINOUE (hinoue@im.wustl.edu)
High quality sequence ston: 453

FEATURES	Source
high quality sequence stop	100
Location/Qualifiers	1 572

```
1. 3/3
source
/organism="Homo sapiens"
/ab_006160000.06000
```

```

/ad_xref=taxon:9006
/clone_lib="HR85 islet"
/clone_type="Bacterial Artificial Chromosome"

```

```
/tissue_type="Purified pancreatic islet"  
/lab_host="DH10B"
```

```

/note="Organ: Pancreas; Vector: pRlucscript SK(-); Site_
NotI; Site_2: XhoI; cDNA made by cllgo-ct priming.

```

Size-selected on agarose gel. Average insert size ~1kb. XhoI site was destroyed after directional cloning.

Amplified once. Contact Information: Hiroshi Inoue, MD, Metabolism Div. (Alan Permitt Lab), Washington University

School of Medicine, Box 8127, 650 South Euclid Ave., St. Louis, MO 63110, E-mail: hinoue@imgate.wustl.edu, Tel:

BASE COUNT	314-362-1916, Fax: 314-747-2692."		
183 a	120 c	71 g	199 t

ORIGIN

Query Match	67.28;	Score 451.4;	DB 13;	Length 573;
Best Local Similarity	99.88;	Pred. No. 3.1e-56;		

Matches 452; Conservative 0; Mismatches 1; Indels 0; Gaps 0

156 TCAGATAAAGGAGATGGGTGAGATGCATAGGAACTCAATGCA¹TAAC²TATATTA³TTT⁴GAA⁵ 215

Db 460 TGAGATAAAGGAGATGGGTGAGATGCATAGGAACTCAATGCATACCTATATATATTGAA 401

QY 216 GATTATAGAGAGGAAATAGCAATATGACACAAATGTGTGCGTGGAGC 275
 |||||||
 Db 400 GATTATAGAGAGGAAATAGCAATATGACACAAATGTGTGCGTGGAGC 341
 |||||||
 QY 276 AAGACATCTTTGAAGTCATGAGTTGTGTAGTTTACATCATATTTGTAATAGTGA 335
 |||||||
 Db 340 AAGACATCTTTGAAGTCATGAGTTGTGTAGTTTACATCATATTTGTAATAGTGA 281
 |||||||
 QY 336 CCGTACTCAAAATATAGAGGCTTGAACGCTTACCAATCTGAAATTTGACACA 395
 |||||||
 Db 280 CCGTACTCAAAATATAGAGGCTTGAACGCTTACCAATCTGAAATTTGACACA 221
 |||||||
 QY 396 AGTGTCTTATATATGACATCTATATGTAATATCCAGAACTTGACCTCATCTTAAAT 455
 |||||||
 Db 220 AGTGTCTTATATATGACATCTATATGTAATATCCAGAACTTGACCTCATCTTAAAT 161
 |||||||
 QY 456 ATTATGTGTACATCTCAATATGTGTGATTAATATATCTTCCACAGTAATCTGAAAA 515
 |||||||
 Db 160 ATTATGTGTACATCTCAATATGTGTGATTAATATATCTTCCACAGTAATCTGAAAA 101
 |||||||
 QY 516 CTGATTTGATGTAAGAGCGCTTCTATTTCTGAGTCTTACATCATCTTTT 575
 |||||||
 Db 100 CTGATTTGATGTAAGAGCGCTTCTATTTCTGAGTCTTACATCATCTTTT 41
 |||||||
 QY 576 TATGAGCTATGAATAAATCATTTTAACTGAA 608
 |||||||
 Db 40 TATGAGCTATGAATAAATCATTTTAACTGAA 8
 |||||||

RESULT 5
 BM272151/c
 LOCUS 592 bp mRNA linear EST 20-DEC-2001
 DEFINITION 1g39g02.x1 HR85 1stlet Homo sapiens cDNA 3' similar to SW:G332.HUMAN
 P16422 MAJOR GASTROINTESTINAL TUMOR-ASSOCIATED PROTEIN GAV33-2
 PRECURSOR: mRNA sequence.

ACCESSION BM272151
 VERSION BM272151
 KEYWORDS GI:17965429
 SOURCE EST.
 ORGANISM human.

REFERENCE
 AUTHORS Homo sapiens
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 592)
 Melton,D., Brown,J., Kenly,G., Pernutt,A., Lee,C., Kaestner,K.,
 Lemishka,I., Scease,M., Brestelli,J., Gradwohl,G., Clifton,S.,
 Hillier,L., Marra,M., Pepe,D., Wylie,T., Martin,J., Bistain,A.,
 Schmitt,A., Theising,B., Riltter,E., Ronko,I., Bennett,V., Cardenas
 , M., Gibbons,M., McCann,R., Cole,R., Tsagarishvili,R., Williams,T.,
 Jackson,Y., and Bowers,Y.
 Endocrine Pancreas Consortium
 Unpublished (2000)

TITLE
 JOURNAL
 COMMENT Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
 Endocrine Pancreas Consortium
 Harvard University, Howard Hughes Medical Institute
 Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
 MA 02138
 Tel: 617-495-1812
 Fax: 617-495-8557
 Email: dmelton@bidh.harvard.edu

Library was constructed by Dr. Hiroshi Inoue DNA sequencing by:
 Washington University Genome Sequencing Center For information on
 obtaining a clone please contact: Dr. Hiroshi Inoue
 (hinoue@um.wustl.edu)
 Seq primer: -400P from Gibco
 High quality sequence stop: 448.
 Location/Qualifiers

FEATURES
 source
 1. 592

/organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="HR85 1stlet"
 /tissue_type="Purified pancreatic 1stlet"
 /lab_host="DH10B"
 /note="Organ: Pancreas; Vector: pbluescript SK(-); Site_1:"

NOTI: Site_2: XhoI; cDNA made by oligo-dT priming.
 Size-selected on agarose gel. Average insert size ~1kb. 5'
 XhoI site was destroyed after directional cloning.
 Amplified once. Contact information: Hiroshi Inoue, MD,
 Metabolism Div. (Alan.Pernutt Lab), Washington University
 School of Medicine, Box 8127, 660 South Euclid Ave., St.
 Louis, MO 63110. E-mail: hinoue@imgate.wustl.edu, Tel:
 314-362-1916, Fax: 314-747-2692.

BASE COUNT 188 a 124 c 77 g 203 t
 ORIGIN

Query Match 67.2%; Score 451.4; DB 13; Length 592;
 Best Local Similarity 99.8%; Pred. No. 3e-56;
 Matches 452; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 156 TCAGATTAAGAGAGTGGTGCATGATGATGAGGAACTCAATGATATATATTTGAA 215
 |||||||
 Db 455 TCAGATTAAGAGAGTGGTGCATGATGATGAGGAACTCAATGATATATATTTGAA 396
 |||||||
 QY 216 GATTATAGAGAGGAAATAGCAATATGACACAAATATGATGCGTGGAGC 275
 |||||||
 Db 395 GATTATAGAGAGGAAATAGCAATATGACACAAATATGATGCGTGGAGC 336
 |||||||
 QY 276 AAGACATCTTTGAAGTCATGAGTTGTGTAGTTTACATCATATTTGTAATAGTGA 335
 |||||||
 Db 335 AAGACATCTTTGAAGTCATGAGTTGTGTAGTTTACATCATATTTGTAATAGTGA 276
 |||||||
 QY 336 CCGTACTCAAAATATAGAGGCTTGAACGCTTACCAATCTGAAATTTGACACA 395
 |||||||
 Db 275 CCGTACTCAAAATATAGAGGCTTGAACGCTTACCAATCTGAAATTTGACACA 216
 |||||||
 QY 396 AGTGTCTTATATATGACATCTATATGTAATATCCAGAACTTGACCTCATCTTAAAT 455
 |||||||
 Db 215 AGTGTCTTATATATGACATCTATATGTAATATCCAGAACTTGACCTCATCTTAAAT 156
 |||||||
 QY 456 ATTATGTGTACATCTCAATATGTGTGATTAATATATCTTCCACAGTAATCTGAAAA 515
 |||||||
 Db 155 ATTATGTGTACATCTCAATATGTGTGATTAATATATCTTCCACAGTAATCTGAAAA 96
 |||||||
 QY 516 CTGATTTGATGTAAGAGCGCTTCTATTTCTGAGTCTTACATCATCTTTT 575
 |||||||
 Db 95 CTGATTTGATGTAAGAGCGCTTCTATTTCTGAGTCTTACATCATCTTTT 36
 |||||||
 QY 576 TATGAGCTATGAATAAATCATTTTAACTGAA 608
 |||||||
 Db 35 TATGAGCTATGAATAAATCATTTTAACTGAA 3

RESULT 6
 AM473340/c
 LOCUS 604 bp mRNA linear EST 24-FEB-2000
 DEFINITION hb03h06.x1 NCI-GAP_Gas4 Homo sapiens cDNA clone IMAGE:2882171 3'
 similar to gb:X14758 MAJOR GASTROINTESTINAL TUMOR-ASSOCIATED
 PROTEIN GAV33-2 (HUMAN); mRNA sequence.

ACCESSION AM473340
 VERSION AM473340
 KEYWORDS GI:7043446
 SOURCE EST.
 ORGANISM human.

REFERENCE
 AUTHORS Homo sapiens
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 604)

TITLE
 JOURNAL
 COMMENT NCI-GAP http://www.ncbi.nlm.nih.gov/ncigap.
 Tumor Gene Index
 Unpublished (1997)
 Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov

Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
 Emmert-Buck, M.D., Ph.D.
 CDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center

AI955311/c		627 bp	mRNA	Linear	EST 09-MAR-2000
LOCUS	AI955311				
DEFINITION	wX2ell..x1 NCI-CGAP_Plit1 Homo sapiens cDNA clone IMAGE:2545388 3'				
	similar to g9:xl4758 MAJOR GASTROINTESTINAL TUMOR-ASSOCIATED PROTEIN GA735-2 (HUMAN);, MRA version.				
ACCESSION	AI955311				
VERSION	AI955311.1	GI:5747621			
KEYWORDS	EST.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
AUTHORS	1 (bases 1 to 627)				
TITLE	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap. National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index				
JOURNAL	Unpublished (1997)				
COMMENT	Contact: Robert Strausberg, Ph.D. Email: cgapbs@emall.nih.gov Tissue Procurement: Chris Moskalko, M.D., Ph.D., Michael R. Emmert-Buck, M.D., CDNA Library Preparation: Life Technologies, Inc. CDNA Library Arrayed by: Christa Prange, The I.M.A.G.E. Consortium DNA Sequencing by: Washington University Genome Sequencing Center Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LNLN at: www.bio.lnl.gov/bdrr/image/image.html Insert length: 1680 Std Error: 0.00 Seq primer: -40UP from Gibco High quality sequence stop: 410. Location/Qualifiers 1..627 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE:2545388" /clone_lib="NCI CGAP_Plit1" /rseq_type="four pooled pituitary adenomas" /lab_host="DH10B" /note="Organ: brain; Vector: pCMV-SPORT6; Site_1: SalI; site_2: NotI; Cloned unidirectionally. Primer: Oligo dt. Library constructed by Life Technologies." BASE COUNT 201 a 131 c 79 g 213 t 3 others ORIGIN Query Match 67.2%; Score 451.4; DB 9; Length 627; Best Local Similarity 99.8%; Pred. No. 2.9e-56; Matches 452; Conservative 0; Mismatches 1; Indels 0; Gaps 0;				
Oy	156 TCAGATTAAGAAGGATGGGAGATGCATCGATGGAGACTCAATGCATACTATATATTGA	215			
Db	456 TGAGATAAAGAGANTOGTGACATCATGTGGAGACATCATATATTAATTGGA	397			
Oy	216 GATTATAGAAGAGGGAATAGCAAATGACACAATTCAGAAATGTGTGGTGGAGC	275			
Db	396 GATTATAGAAGAGGGAATAGCAAATGACACAATTCAGAAATGTGTGGTGGAGC	337			
Oy	276 AAGACATCTTTGAAGTCAATGATTTGTTAGATTAAATATATTTGTAATAGGANA	335			
Db	336 AAGACATCTTTGAAGTCAATGATTTGTTAGATTAAATATATTTGTAATAGGANA	277			
Oy	336 CCTGTACTCAAAATATAMGACCTTGAACCTGGCTTTACCAGCTTGAATTTGACACA	395			
Db	276 CCGTGTACTCAAAATATAMGACCTTGAACCTGGCTTTACCAGCTTGAATTTGACACA	217			
Oy	396 AGTGCTTATATATGCAAGATCTAATCTAAATCCAGAACTTGAGACCTCATGTTAAAT	455			
Db	216 AGTGCTTATATATGCAAGATCTAATCTAAATCCAGAACTTGAGACCTCATGTTAAAT	157			
Oy	456 ATTATATGTTAACATTCGAATGTGTGTCATTTAAATATCTTCCACAGTAATCTGAAAA	515			
Db	156 ATTATATGTTAACATTCGAATGTGTGTCATTTAAATATCTTCCACAGTAATCTGAAAA	97			
Oy	516 CTGATTTGATGTAAGAGCTGCCCTTTCTATTTACTGAGCTCTTGATACATACACTTTTT	575			

Db	96	CTGATTGTGATGTAAGACCTGCCTTTCATTTACTTGAAGTGAGCTTGACATCATCACTTTTT	
QY	576	TATGAGCTATGAATAAACAATTTTAACGTAA 608	
Db	36	TATGAGCTATGAATAAATAACATTTTAACTGAA 4	
RESULT 9			
AI978632/c			
LOCUS	AI978632	675 bp	mRNA linear EST 08-MAR-2000
DEFINITION	w57e02.x1 NCI-CGAP_Uci Homo sapiens cDNA clone IMAGE:2491802 3'		
	similar to gb:X14758 MAJOR GASTROINTESTINAL TUMOR-ASSOCIATED		
	PROTEIN GA733-2 (HUMAN);, mRNA sequence.		
ACCESSION	AI978632		
VERSION	AI978632.1 GI:5803662		
KEYWORDS	EST.		
SOURCE	Homo sapiens		
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.		
REFERENCE	1 (bases 1 to 675)		
AUTHORS	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.		
TITLE	National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index Unpublished (1997)		
JOURNAL	Contact: Robert Strausberg, Ph.D. Email: cgapbs@mail.nih.gov Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D. CDNA Library Preparation: Life Technologies, Inc. cDNA Library Arrayed by: Greg Lennon, Ph.D. DNA Sequencing by: Washington University Genome Sequencing Center Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: www.bio.lnl.gov/db/rp/image/image.html Insert length: 1133 Std Error: 0.00 Seq primer: -40bp from Glbco High quality sequence stop: 396.		
FEATURES	Location/Qualifiers		
source	1..675		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/clone="IMAGE:2491802"		
	/clone_1lb="NCI CGAP utl1"		
	/tissue_type="well-differentiated endometrial adenocarcinoma, 7 pooled tumors"		
	/lab_host="DH10B"		
	/note="Organ: uterus; Vector: PCMV-SPOK6; Site:1: SalI; Site:2: NotI; Cloned unidirectionally. Primer: Oligo dt. Average insert size 1.75 kb. Life technologies catalog #: 11538-014"		
BASE COUNT	210 a 141 c 89 g 234 t 1 others		
ORIGIN			
Query Match	67.2%; Score 451.4; DB 9; Length 675;		
Best Local Similarity	99.8%; Pred. No. 2.8e-56;		
Matches 452; Conservative 0; Mismatches 1; Indels 0; Gaps 0;			
QY	156	TCAGTAAAGGAGATGGCTGAGATGATCATGAGGAAGCAATGCATATATATATATTGAA 215	
Db	460	TCAGTAAAGGAGATGGGAGATGATCATAAGGAACTCATGATATATATATATTGAA 401	
QY	216	GATTATAGAAGAGGAATATACCAATGAGACCAAAATTCACAAATGCTGTGCGTGGAGC 275	
Db	400	GATTATAGAAGAGGAATATACCAATGAGACCAAAATTCACAAATGCTGTGCGTGGAGC 341	
QY	276	AAGACATCTTTGAAGCTCATGAGTTGTTAGTTAACATCATATATTTGTATATAGTAAA 335	
Db	340	AAGACATCTTTGAAGCTCATGAGTTGTTAGTTAACATCATATATTTGTATATAGTAAA 281	
QY	336	CCTGATCCAAAAATATACACAGCTGAACCTGACCTTACCATCTTGAATTTGACACA 395	

Db 280 CCTGACTCAAAATATTAAGACGCTTGAAACGTGGCTTTACCAATCTTGAATTTGACACCA 221

Qy 396 AGTGTCTATATATGACAGATCTAATGTAAATCCAGAACTGGATCCATGCTTAAAT 455

Db 220 AGTGTCTATATATGACAGATCTAATGTAAATCCAGAACTGGATCCATGCTTAAAT 161

Qy 456 ATTTATGTGTACATTCAAATGTGTGCTTAAATGTCTCCACAGTAAATCTGAAAA 515

Db 160 ATTTATGTGTACATTCAAATGTGTGCTTAAATGTCTCCACAGTAAATCTGAAAA 101

Qy 516 CTGATTTGTGATGAAAGCTGCTTCTATTTACTGTGATGATCAATCACTTTT 575

Db 100 CTGATTTGTGATGAAAGCTGCTTCTATTTACTGTGATGATCAATCACTTTT 41

Qy 576 TATGAGCTATGAAATTAACATTTTAACTGAA 608

Db 40 TATGAGCTATGAAATTAACATTTTAACTGAA 8

RESULT 10

BM981479/c 709 bp mRNA linear EST 21-MAR-2002

LOCUS UI-CF-EN1-adh-h-02-0-UI.s1 UI-CF-EN1 Homo sapiens cDNA clone

DEFINITION UI-CF-EN1-adh-h-02-0-UI 3', mRNA sequence.

ACCESSION BM981479

VERSION BM981479.1 GI:19604010

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 709)

AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.

TITLE Normalization and subtraction: two approaches to facilitate gene discovery

JOURNAL Genome Res. 6 (9), 791-806 (1996)

MEDLINE 97044477

COMMENT Contact: McCray, PB

McCray Lab

2024 University of Iowa Med Labs., Iowa City, IA 52242, USA

Tel: 319 356 4866

Fax: 319 356 7171

Email: paul-mccray@uiowa.edu

Tissue Procurement: Dr. M. J. Welsh, University of Iowa

cDNA Library preparation: Dr. M. Bento Soares, University of Iowa

cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa

Genetics (www.resgen.com).

The following repetitive elements were found in this cDNA

sequence: 505-573, >(CAA)n#Simple_repeat

Seq primer: M13 FORWARD

PolyA=yes.

FEATURES

Source Location/Qualifiers

1..709

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="UI-CF-EN1-adh-h-02-0-UI"

/clone_lib="UI-CF-EN1"

/tissue_type="Primary Lung Cystic Fibrosis Epithelial Cells"

/dev_stage="Adult"

/lab_host="DH10B (Life Technologies) (T1 phage resistant)"

/note="Organ: Lung; Vector: pRT3-Pac (Pharmacia) with a modified polylinker; Site_1: EcoR I; Site_2: Not I; UI-CF-EN1 is a normalized cDNA library containing the following tissue(s): Primary Lung Cystic Fibrosis Epithelial Cells. The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not

I, and cloned directionally into pRT3-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (df)18 tail. The sequence tag for this library is CTGCTCAGGT.

TAG_Lib-UI-CF-EN1

TAG_Tissue-Human Lung Epithelial Cell Lines untreated LPS

6hr to LPS 24h

TAG_SEQ-CTGCTCAGGT"

BASE COUNT 221 a 146 c 91 g 251 t

ORIGIN

Query Match 67.2%; Score 451.4; DB 14; Length 709;

Best Local Similarity 99.8%; Pred. No. 2.7e-56;

Matches 452; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 156 TCAGATAAAGAGATGGGTGATGCATAGGAAGCAATGCATATATTAATTGAA 215

Db 469 TGAGATTAAGAGATGGGTGATGCATAGGAAGCAATGCATATATTAATTGAA 410

Qy 216 GATTATAGAGAGAGGAATAGCAATGACACAAATTTGATGCGTGGAGC 275

Db 409 GATTATAGAGAGAGGAATAGCAATGACACAAATTTGATGCGTGGAGC 350

Qy 276 AAGACATCTTTGAGGTCATGAGTTGTAGTTTAACATCATATATTTGTAATAGTGANA 335

Db 349 AAGACATCTTTGAGGTCATGAGTTGTAGTTTAACATCATATATTTGTAATAGTGANA 290

Qy 336 CCGTACATCAAAATATTAAGCAAGCTGGAAGCTTTTACCAATCTGAAATTTGACACCA 365

Db 289 CCGTACATCAAAATATTAAGCAAGCTGGAAGCTTTTACCAATCTGAAATTTGACACCA 230

Qy 396 AGTGTCTATATATGACAGATCTAATGTAAATCCAGAACTGGATCCATGCTTAAAT 455

Db 220 AGTGTCTATATATGACAGATCTAATGTAAATCCAGAACTGGATCCATGCTTAAAT 170

Qy 456 ATTTATGTGTACATTCAAATGTGTGCTTAAATGTCTCCACAGTAAATCTGAAAA 515

Db 160 ATTTATGTGTACATTCAAATGTGTGCTTAAATGTCTCCACAGTAAATCTGAAAA 110

Qy 516 CTGATTTGTGATGAAAGCTGCTTCTATTTACTGTGATGATCAATCACTTTT 575

Db 100 CTGATTTGTGATGAAAGCTGCTTCTATTTACTGTGATGATCAATCACTTTT 50

Qy 576 TATGAGCTATGAAATTAACATTTTAACTGAA 608

Db 49 TATGAGCTATGAAATTAACATTTTAACTGAA 17

RESULT 11

A1571882/c 768 bp mRNA linear EST 12-MAY-1999

LOCUS A1571882

DEFINITION to20f08.x1 NCI_CGAP_Ut2 Homo sapiens cDNA clone IMAGE:2179623 3'

Similar to gb:U14758 MAJOR GASTROINTESTINAL TUMOR-ASSOCIATED PROTEIN GA733-2 (HUMAN); mRNA sequence.

ACCESSION A1571882

VERSION A1571882.1 GI:4535256

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 768)

AUTHORS NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.

TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index

JOURNAL Unpublished (1997)

COMMENT Contact: Robert Strausberg, Ph.D.

Email: cga@bbs.femail.nih.gov

Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: Greg Lennon, Ph.D.

DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/dbp/image/image.html

Insert Length: 1567 Std Error: 0.00
Seq primer: -40UP from Glibco

High quality sequence stop: 414
POLYA-No.

FEATURES

source

Location/Qualifiers
1..768
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2179623"
/clone_1lb="NCI-CGAP_U2"
/tissue_type="moderately-differentiated endometrial
adenocarcinoma, 3 pooled tumors"
/lab_host="DH10B"
/note="Organ: uterus; Vector: pCMV-Sport6; Site_1: SalI;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.85 kb. Life Technologies catalog #:
11539-012"

BASE COUNT 243 a 160 c 102 g 260 t 3 others
ORIGIN

Query Match 67.2%; Score 451.4; DB 9%; Length 768;
Best Local Similarity 99.8%; Pred. No. 2,6e-56;
Matches 452; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
OY 156 TCAGATTAAGGAGATGGGTCATGATGAGGACATGCTATCTATATATTGAA 215
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 458 TGAGATAAGGAGATGGGTCATGATGAGGACATGCTATCTATATATTGAA 399
OY 216 GATTATAGAGAAGGAGATGCAATGACACAAATGCTGCTGGGAGC 275
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 398 GATTATAGAGAAGGAGATGCAATGACACAAATGCTGCTGGGAGC 339
OY 276 AAGACATCTTTGAAGTCATGATGTTAGTTAAATCATATATTTGTAATGAAA 335
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 338 AAGACATCTTTGAAGTCATGATGTTAGTTAAATCATATATTTGTAATGAAA 279
OY 336 CCGTACTCAAAATATAGACGCTTGAACCTGCTTACCAATCTGAATTTGACCACA 395
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 278 CCGTACTCAAAATATAGACGCTTGAACCTGCTTACCAATCTGAAATTTGACCACA 219
OY 396 AGTGTCTTATATGAGATCTAATGTAAATCCAGAACTGGACTCCATCGTTAAAT 455
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 218 AGTGTCTTATATGAGATCTAATGTAAATCCAGAACTGGACTCCATCGTTAAAT 159
OY 456 ATTATGTGTACATTCATGCTGCTATTAATATGCTTCCACAGTAAATCTGAAAA 515
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 158 ATTATGTGTACATTCATGCTGCTATTAATATGCTTCCACAGTAAATCTGAAAA 99
OY 516 CCGATTTGATGAGAAAGCTGCTTCTATTTACTGAGCTGTGATACATACATCTTT 575
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 98 CCGATTTGATGAGAAAGCTGCTTCTATTTACTGAGCTGTGATACATACATCTTT 39
OY 576 TATGAGCTATGAATTAACATTTTAACTGAA 608
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 38 TATGAGCTATGAATTAACATTTTAACTGAA 6
```

RESULT 12

AA502398/c 513 bp mRNA linear EST 19-AUG-1997
LOCUS ne27h08.s1 NCI-CGAP_C03 Homo sapiens CDNA clone IMAGE:898623 3'
DEFINITION similar to gb:U14758 MAJOR GASTROINTESTINAL TUMOR-ASSOCIATED
PROTEIN GAT33-2 (HUMAN);, mRNA sequence.

ACCESSION AA502398
VERSION AA502398.1 GI:2237365

KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
1 (bases 1 to 513)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Elias Campo, M.D., Michael R. Emmert-Buck, M.D.
Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arraying: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/dbp/image/image.html
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 352.

FEATURES

source

Location/Qualifiers
1..513
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:898623"
/clone_1lb="NCI-CGAP_C03"
/sex="pooled"
/tissue_type="colon"
/lab_host="DH10B"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker. Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was prepared from 12 pooled bulk tumor samples and primed
with a Not I - oligo(dT) primer. Double-stranded cDNA was
ligated to Eco RI adaptors (Pharmacia), digested with Not
I and cloned into the Not I and Eco RI sites of the
modified pT73 vector. Library went through one round of
normalization."

BASE COUNT 161 a 102 c 69 g 181 t
ORIGIN

Query Match 67.0%; Score 450.4; DB 9%; Length 513;
Best Local Similarity 99.8%; Pred. No. 4,6e-56;
Matches 451; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
OY 156 TCAGATTAAGGAGATGGGTCATGATGAGGACATGCTATCTATATATTGAA 215
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 452 TGAGATAAGGAGATGGGTCATGATGAGGACATGCTATCTATATATTGAA 393
OY 216 GATTATAGAGAAGGAGATGCAATGACACAAATGCTGCTGGGAGC 275
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 392 GATTATAGAGAAGGAGATGCAATGACACAAATGCTGCTGGGAGC 333
OY 276 AAGACATCTTTGAAGTCATGATGTTAGTTAAATCATATATTTGTAATGAAA 335
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 332 AAGACATCTTTGAAGTCATGATGTTAGTTAAATCATATATTTGTAATGAAA 273
OY 336 CCGTACTCAAAATATAGACGCTTGAACCTGCTTACCAATCTGAATTTGACCACA 395
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 272 CCGTACTCAAAATATAGACGCTTGAACCTGCTTACCAATCTGAATTTGACCACA 213
OY 396 AGTGTCTTATATGAGATCTAATGTAAATCCAGAACTGGACTCCATCGTTAAAT 455
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 212 AGTGTCTTATATGAGATCTAATGTAAATCCAGAACTGGACTCCATCGTTAAAT 153
OY 456 ATTATGTGTACATTCATGCTGCTATTAATATGCTTCCACAGTAAATCTGAAAA 515
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 152 ATTATGTGTACATTCATGCTGCTATTAATATGCTTCCACAGTAAATCTGAAAA 93
OY 516 CCGATTTGATGAGAAAGCTGCTTCTATTTACTGAGCTGTGATACATACATCTTT 575
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 92 CCGATTTGATGAGAAAGCTGCTTCTATTTACTGAGCTGTGATACATACATCTTT 33
OY 576 TATGAGCTATGAATTAACATTTTAACTGAA 607
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 32 TATGAGCTATGAATTAACATTTTAACTGAA 1
```

RESULT	13
LOCUS	A1038566/c
DEFINITION	A1038566 471 bp mRNA linear EST 28-AUG-1998
ACCESSION	U034B10.s1 Soares-total_fetus_ND2HF8_9w Homo sapiens cDNA clone
KEYWORDS	IMAGE:1658203 3' similar to gb.X14758 MAJOR GASTROINTESTINAL TUMOR-ASSOCIATED PROTEIN GA733-2 (HUMAN); mRNA sequence.
ORGANISM	A1038566
SOURCE	A1038566 GI:3277760
REFERENCE	EST.
AUTHORS	human.
TITLE	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo. 1 (bases 1 to 471)
JOURNAL	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap. National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index Unpublished (1997)
COMMENT	Contact: Robert Strausberg, Ph.D. Email: cgapsb@emall.nih.gov This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.lnl.gov) for further information. Insert Length: 1379 Std Error: 0.00 Seq primer: -40m13 fwd. ET from Amersham High quality sequence stop: 367.
FEATURES	location/Oualifiers 1..471 /organism="Homo sapiens" /db_xref="taxon:9606" /clone_image="1658203" /clone_id="Soares_total_fetus_ND2HF8_9w" /dev_stage="-8-9 weeks" /lab_host="DH10B"
source	/note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker. Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was prepared from mRNA obtained from pooled 8-9 week (total) fetus material with a Not I - oligo(dT) primer [5' TGTTACAATCTGAAGTGAGGGCCGTAAATTCTTTTTTTTTTTT 3'] Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization, and was constructed by Bento Soares and M. Fatima Bonaldo."
BASE COUNT	148 a 87 c 61 g 175 t
ORIGIN	
Query Match	66.9%; Score 449.8; DB 9; Length 471;
Best Local Similarity	99.6%; Pred. No. 5.9e-56;
Matches 451; Conservative	0; Mismatches 2; Indels 0; Gaps 0;
OY	156 TCAGTATAAAGAGATGGGTGAGATGCATFAGGAGCACTCATGCTATATTAATTGAA 215
Db	1 TT 1
460	TGAGTATAAAGAGATGGGTGAGATGCATFAGGAGCACTCATGCTATATTAATTGAA 401
OY	216 GATTATAGAAGAAGGAATAGCAAATGAGACAAATATACAAATGTGTGCGTGGAGC 275
Db	1 TT 1
400	GATTATAGAAGAAGGAATAGCAAATGAGACAAATATACAAATGTGTGCGTGGAGC 341
OY	276 AAGACATCTTGAAGTCATGAGTTGTGTACTTTAACATCATATATTTGTAATAGTAAA 335
Db	1 TT 1
340	AAGACATCTTGAAGTCATGAGTTGTGTACTTTAACATCATATATTTGTAATAGTAAA 281
OY	336 CCTGTACCAAAATATATAGCAGCTTGAACGTGGCTTTACCAATCTTGAAATTTGACCACA 395
Db	1 TT 1
280	CCTGTACCAAAATATATAGCAGCTTGAACGTGGCTTTACCAATCTTGAAATTTGACCACA 221
OY	396 AGTGTCTTATATATGAGATCTAATGTMAATCCAGAACTTGGACTCCATGTTAAATT 455
Db	1 TT 1
220	AGTGTCTTATATATGAGATCTAATGTMAATCCAGAACTTGGACTCCATGTTAAATT 161
OY	456 ATTATGTGTACATTCAAATGTGCTTAAATATGCTTCCACAGTAAATTCGAAAAA 515

QY	DB	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
QY 516	DB 100	BM510869	1j44c04.x1 Human insulinoma Homo sapiens cDNA clone IMAGE:5633670 3', similar to SW:G732_HUMAN P16422 MAJOR GASTROINTESTINAL TUMOR-ASSOCIATED PROTEIN GA733-2 PRECURSOR ;, mRNA sequence.	BM510869	BM510869.1	GI:18682012	EST.	Homo sapiens human.	Mammalia; Eutheria; Primates; Carnivora; Homiidae; Homo.	1 (bases 1 to 461)	Melton, D., Brown, J., Kenty, G., Permutt, A., Lee, C., Kastner, K., Lemishka, I., Searce, M., Brestelli, J., Gradwohl, G., Clifton, S., Hillier, L., Marra, M., Page, D., Wylie, T., Martin, J., Blistain, A., Schmitt, A., Theising, B., Ritter, E., Ronko, J., Bennett, J., Cardenas, M., Gibbons, M., McCann, R., Cole, R., Tsagaratishvili, R., Williams, T., Jackson, Y., and Bowers, Y.	Endocrine Pancreas Consortium Unpublished (2000)	Contract: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue Harvard University, Howard Hughes Medical Institute Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge, MA 02138 Tel: 617-495-1812 Fax: 617-495-8557 Email: dmelton@hmp.harvard.edu Library was constructed by Dr. J. Ferrer in vivo mass-excised to pBluescript SK- by Dr. H. Inoue DNA sequencing by: Washington University Genome Sequencing Center For information on obtaining a clone please contact: Dr. Hiroshi Inoue (hinoue@im.wustl.edu) Seq primer: -400p from Gibco.
QY 576	DB 40	TATGAGCTATGAAATTAACATTTTAACGTGA 608						TATGAGCTATGAAATTAACATTTTAACGTGA 8					
FEATURES	Source	1. 481	/organism="Homo sapiens"	/db_xref="taxon:9606"	/clone="IMAGE:5633670"	/clone_id="Human insulinoma"	/tissue_type="insulinoma"	/lab_host="DH10B (phage-resistant)"	/note="Organ: pancreas; Vector: pBluescript SK-; Site_1: XhoI; Site_2: EcoRI; Constructed with lambda ZAPII system (Stratagene) by Dr. J. Ferrer, in vivo mass-excised to pBluescript SK- by Dr. H. Inoue following the Washington University protocol (http://genome.wustl.edu/est/lambda-protocol.shtml). Please contact Hiroshi Inoue, MD/PhD for further information on this library (Metabolism Division, Permutt Laboratory, Washington University School of Medicine, Box 8127, 660 S Euclid Ave, St. Louis, MO 63110). Note: this is a Washington University Pancreas EST project library."				
BASE COUNT	146 a	88 c	63 g	184 t									
ORIGIN	Query Match	66.9%	Score 449.8	DB 13	Length 481								
	Best Local Similarity	99.6%	Pred. No. 5.8e-56										
	Matches 451	Conservative	0	Mismatches	2	Indels	0	Gaps	0				

NAME/KEY: EXON
LOCATION: 3871-4072

1	OTHER INFORMATION: exon 2	
2	NAME/KEY: exon	
3	LOCATION: 5552..5633	
4	OTHER INFORMATION: exon 3	
5	NAME/KEY: exon	
6	LOCATION: 5758..5880	
7	OTHER INFORMATION: exon 4	
8	NAME/KEY: exon	
9	LOCATION: 5996..6099	
10	OTHER INFORMATION: exon 5	
11	NAME/KEY: exon	
12	LOCATION: 6349..6509	
13	OTHER INFORMATION: exon 6	
14	NAME/KEY: exon	
15	LOCATION: 7379..7522	
16	OTHER INFORMATION: exon 7	
17	NAME/KEY: exon	
18	LOCATION: 8645..8854	
19	OTHER INFORMATION: exon 8	
20	NAME/KEY: exon	
21	LOCATION: 12254..12340	
22	OTHER INFORMATION: exon 9	
23	NAME/KEY: exon	
24	LOCATION: 12854..13023	
25	OTHER INFORMATION: exon 10	
26	NAME/KEY: exon	
27	LOCATION: 13308..13429	
28	OTHER INFORMATION: exon 11	
29	NAME/KEY: exon	
30	LOCATION: 16567..16667	
31	OTHER INFORMATION: exon 12	
32	NAME/KEY: exon	
33	LOCATION: 16775..16945	
34	OTHER INFORMATION: exon 13	
35	NAME/KEY: exon	
36	LOCATION: 17063..17554	
37	OTHER INFORMATION: exon 14	
38	NAME/KEY: misc feature	
39	LOCATION: 17555..20674	
40	OTHER INFORMATION: 3 regulatory region	
41	NAME/KEY: allele	
42	LOCATION: 1128	
43	OTHER INFORMATION: 10-508-191	: polymorphic base C or T
44	NAME/KEY: allele	
45	LOCATION: 1182	
46	OTHER INFORMATION: 10-508-245	: polymorphic base C or T
47	NAME/KEY: allele	
48	LOCATION: 1559	
49	OTHER INFORMATION: 10-509-284	: polymorphic base C or T
50	NAME/KEY: allele	
51	LOCATION: 1570	
52	OTHER INFORMATION: 10-509-295	: deletion of C
53	NAME/KEY: allele	
54	LOCATION: 1827	
55	OTHER INFORMATION: 10-510-173	: variable motif AATTGA or TTTTGT
56	NAME/KEY: allele	
57	LOCATION: 2048	
58	OTHER INFORMATION: 10-511-62	: polymorphic base C or T
59	NAME/KEY: allele	
60	LOCATION: 2323	
61	OTHER INFORMATION: 10-511-337	: insertion of T
62	NAME/KEY: allele	
63	LOCATION: 2341	
64	OTHER INFORMATION: 10-512-36	: polymorphic base G or C
65	NAME/KEY: allele	
66	LOCATION: 2623	
67	OTHER INFORMATION: 10-512-318	: polymorphic base A or G
68	NAME/KEY: allele	
69	LOCATION: 2832	
70	OTHER INFORMATION: 10-513-250	: polymorphic base A or G
71	NAME/KEY: allele	
72	LOCATION: 2844	
73	OTHER INFORMATION: 10-513-265	: polymorphic base C or T

NAME/KEY: allele
LOCATION: 2934 : polymorphic base A or G
OTHER INFORMATION: 10-513-352 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 2947 : polymorphic base A or G
OTHER INFORMATION: 10-513-365 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 3802 : polymorphic base A or G
OTHER INFORMATION: 12-206-81 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 4062 : deletion of C
OTHER INFORMATION: 10-343-231 : deletion of C
NAME/KEY: allele
LOCATION: 4088 : polymorphic base C or T
OTHER INFORMATION: 12-206-366 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 4109 : polymorphic base C or T
OTHER INFORMATION: 10-343-278 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 4170 : polymorphic base G or T
OTHER INFORMATION: 10-343-339 : polymorphic base G or T
NAME/KEY: allele
LOCATION: 5903 : polymorphic base A or G
OTHER INFORMATION: 10-346-23 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 6019 : polymorphic base A or G
OTHER INFORMATION: 10-346-141 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 6141 : polymorphic base G or C
OTHER INFORMATION: 10-346-263 : polymorphic base G or C
NAME/KEY: allele
LOCATION: 6183 : polymorphic base C or T
OTHER INFORMATION: 10-346-305 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 6338 : polymorphic base A or G
OTHER INFORMATION: 10-347-74 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 6375 : polymorphic base G or C
OTHER INFORMATION: 10-347-111 : polymorphic base G or C
NAME/KEY: allele
LOCATION: 6429 : polymorphic base C or T
OTHER INFORMATION: 10-347-165 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 6467 : polymorphic base A or G
OTHER INFORMATION: 10-347-203 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 6484 : polymorphic base A or G
OTHER INFORMATION: 10-347-220 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 6534 : polymorphic base A or T
OTHER INFORMATION: 10-347-271 : polymorphic base A or T
NAME/KEY: allele
LOCATION: 6611 : polymorphic base A or G
OTHER INFORMATION: 10-347-348 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 7668 : polymorphic base A or G
OTHER INFORMATION: 10-348-391 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 8608 : polymorphic base C or T
OTHER INFORMATION: 10-349-47 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 8658 : polymorphic base A or G
OTHER INFORMATION: 10-349-97 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 8703 : polymorphic base G or C
OTHER INFORMATION: 10-349-142 : polymorphic base G or C
NAME/KEY: allele
LOCATION: 8777 : deletion of CTG
OTHER INFORMATION: 10-349-216 : deletion of CTG
NAME/KEY: allele
LOCATION: 8785 : polymorphic base G or T
OTHER INFORMATION: 10-349-224 : polymorphic base G or T
NAME/KEY: allele

LOCATION: 8926 : polymorphic base C or T
OTHER INFORMATION: 10-349-368 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 12171 : polymorphic base C or T
OTHER INFORMATION: 10-350-72 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 12429 : polymorphic base C or T
OTHER INFORMATION: 10-350-332 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 13341 : polymorphic base A or G
OTHER INFORMATION: 10-507-170 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 13492 : polymorphic base A or C
OTHER INFORMATION: 10-507-321 : polymorphic base A or C
NAME/KEY: allele
LOCATION: 13524 : polymorphic base C or T
OTHER INFORMATION: 10-507-353 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 13535

Query Match 7.4% Score 50; DB 4; Length 20674;
Best Local Similarity 46.1%; Pred. No. 0.095;
Matches 167; Conservative 0; Mismatches 195; Indels 0; Gaps 0;

QY 303 TTAGTTTACATCATATTTTGTATAGTGAACCTGTACTCAAAATATACGAGCTTGA 362
DB 11103 TTAATTAATTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTT 11162
QY 363 AACTGCTTTACCATCTTGAATTTTGACCACAGTCTTATATATGACATCTAATGT 422
DB 11163 TAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTTAATTT 11222
QY 423 AAAATCCAGACTGTGACCCATCGTAAATTTATTTAGTGAACATCAATGCTGC 482
DB 11223 TAAATTAATTAATTAATTTTAAATTTAATTAATTAATTTAATTTAATTTAATTTA 11282
QY 483 ATTAATATGCTTCCACAGTAAATCTGAATAAAGTATTTGATGAAAGCTGCTTC 542
DB 11283 ACTAATTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTTAATTT 11342
QY 543 TATTTACTTGAGCTTGTACATACATACATCTTTTATGACCTATGAATTAATTAATTTAA 602
DB 11343 TTAATTAATTTAATTTAATTAATTAATTAATTTAATTTAATTTAATTTAATTTAATTTAA 11402
QY 603 ACGAATTTCTTAACCTTGACATTTCAATTTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 662
DB 11403 TTTTAATTTAATTTGCTCAATTAATTAATTTAATTTAATTTAATTTAATTTAATTTAA 11462
QY 663 TT 664
DB 11463 TT 11464

RESULT 4
US-07-867-106-2
Sequence 2, Application US/07867106
Patent No. 5389526
GENERAL INFORMATION:
APPLICANT: Slade, Martin B
APPLICANT: Chang, Andy C M
APPLICANT: Williams, Keith L
TITLE OF INVENTION: Improved Plasmid Vectors for Cellular
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 5389526ris
STREET: One Liberty Place 46th Floor
CITY: Philadelphia
STATE: PA
COUNTRY: USA
ZIP: 19103
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk


```
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/867,106
FILING DATE: 19920625
PRIOR APPLICATION DATA:
APPLICATION NUMBER: AU PJ 7187
APPLICATION NUMBER: PCT/AU90/00530
FILING DATE: 02-NOV-1989
ATTORNEY/AGENT INFORMATION:
NAME: Feeney, Joanne Longo
REGISTRATION NUMBER: 35,134
REFERENCE/DOCKET NUMBER: RICE-0002
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-568-3100
TELEFAX: 215-568-3439
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 5852 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ANTI-SENSE: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 2378..5038
FEATURE:
NAME/KEY: CDS
LOCATION: 2378..5038
US-07-867-106-2
```

Query Match 7.4%; Score 49.6; DB 1; Length 5852;
Best Local Similarity 54.3%; Pred. No. 0.099;
Matches 100; Conservative 0; Mismatches 84; Indels 0; Gaps 0;

```
QY 484 TTAATATGCTTCACAGTAATCTGAAAACGATTTGATTTGAAGCTGCTTCT 543
    ||||| - - - - - ||||| - - - - - ||||| - - - - - |||||
Db 5445 TTAATAAAAAAAAAAAAAAAAAAAAAATCAATATGTTTATGTTTAAAGATTTT 5504

QY 544 ATTACTGAGCTCTTGACATACACTTTTATGAGCTTGAAATTAACATTTTAA 603
    ||||| - - - - - ||||| - - - - - ||||| - - - - - |||||
Db 5505 AATCTCGCAATGATTTTAAATGCAATGATTAATTTTAAAAAACCCTTACA 5564

QY 604 CTGAATTTCTTAACCTTGACATTTCAATTTCTCTCTTTCTTTCTTTCTTT 663
    ||||| - - - - - ||||| - - - - - ||||| - - - - - |||||
Db 5565 TTTTATTTTAAATTCCAATTTTATACATTTTATTTTATTTTATTTTATTT 5624

QY 664 TTTT 667
    ||||
Db 5625 TTTT 5628
```

RESULT 5
US-09-129-112-3/c
Sequence 3, Application US/09129112
Patent No. 6465716
GENERAL INFORMATION:
APPLICANT: Etzler, Marilyn E.
APPLICANT: Murphy, Judith B.
TITLE OF INVENTION: A No. 6465716 Factor binding protein from Legume Roots
FILE REFERENCE: 023070-079810US
CURRENT APPLICATION NUMBER: US/09/129,112
CURRENT FILING DATE: 1998-08-04
PRIOR APPLICATION NUMBER: US 08/907,226
FILING DATE: 1997-08-06
NUMBER OF SEQ ID NOS: 19
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 3
LENGTH: 6265
TYPE: DNA

```
ORGANISM: Dolichos biflorus
FEATURE:
OTHER INFORMATION: genomic sequence of NBP46 (DB46)
NAME/KEY: exon
LOCATION: (633)..(944)
NAME/KEY: intron
LOCATION: (945)..(1022)
NAME/KEY: exon
LOCATION: (1023)..(1151)
NAME/KEY: intron
LOCATION: (1152)..(1559)
NAME/KEY: exon
LOCATION: (1560)..(1616)
NAME/KEY: intron
LOCATION: (1617)..(1697)
NAME/KEY: exon
LOCATION: (1698)..(1790)
US-09-129-112-3
```

Query Match 7.3%; Score 49.2; DB 4; Length 6265;
Best Local Similarity 47.2%; Pred. No. 0.12;
Matches 150; Conservative 0; Mismatches 168; Indels 0; Gaps 0;

```
QY 351 TAAGCAGCTTGAAACGCTTTACCAATCTGAAATTTGACACAGCTGCTTATATG 410
    ||||| - - - - - ||||| - - - - - ||||| - - - - - |||||
Db 3720 TTAAGAGTAAATAATTGACAGAAAAAATACATATGATATTAATTAATTAAT 3661

QY 411 CAGATCTAATGTAAATCCAGAACTTGACGCCATCGTTAAATTTATTTATGTACAT 470
    ||||| - - - - - ||||| - - - - - ||||| - - - - - |||||
Db 3660 AACATTAATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTT 3601

QY 471 TCAATGTGTCATTAATATGCTTCCACAGTAATAATCTGAAAACGATTTGATTTGA 530
    ||||| - - - - - ||||| - - - - - ||||| - - - - - |||||
Db 3600 ATATTTATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 3541

QY 531 AAGTCGCTTCTATTTACTTGAGCTCTTGACATACATTTCTTTTATGAGCTATGAAT 550
    ||||| - - - - - ||||| - - - - - ||||| - - - - - |||||
Db 3540 TAAATTTAAATAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTT 3481

QY 591 AAACATTTTAACTGAATTTCTTAACCTTGACATTTCAATTTCTCTCTTTCTTT 650
    ||||| - - - - - ||||| - - - - - ||||| - - - - - |||||
Db 3480 ATTAATAATTAATTTAACTTTCTCTCTTAATTAATTAATTAATTAATTAAT 3421

QY 651 TCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCT 668
    ||||| - - - - - ||||| - - - - - ||||| - - - - - |||||
Db 3420 TCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCT 3403
```

RESULT 6
US-08-213-419B-3/c
Sequence 3, Application US/08213419B
Patent No. 6333406
GENERAL INFORMATION:
APPLICANT: Inselburg, J. et al.
TITLE OF INVENTION: GENE ENCODING PROTEIN ANTIGENS OF PLASMODIUM FALCIPARUM
FILE REFERENCE: JTI-002CNCPE
CURRENT APPLICATION NUMBER: US/08/213,419B
CURRENT FILING DATE: 1994-03-14
PRIOR APPLICATION NUMBER: US 07/870,506
FILING DATE: 1992-04-17
NUMBER OF SEQ ID NOS: 20
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 3
LENGTH: 6124
TYPE: DNA
ORGANISM: Plasmodium falciparum
FEATURE:
NAME/KEY: CDS
LOCATION: (2407)..(2439)
NAME/KEY: CDS
LOCATION: (2598)..(3404)
NAME/KEY: CDS

```
; LOCATION: (3580)..(3720)
; NAME/KEY: CDS
; LOCATION: (3850)..(5835)
US-08-213-419B-3
```

Query Match	7.28;	Score: 48.6;	DB 4;	Length 6124;
Best Local Similarity	64.98;	Pred. No. 0.16;		
Matches	72;	Conservative	0;	Mismatches 39;
			Indels	0;
			Gaps	0

[illegible]

```

RESULT 7
US-08-973-462-1
: Sequence 1, Application US/08973462B
: Patent No. 6191270
:
GENERAL INFORMATION:
:
APPLICANT: DROTHLE, PIERRE
APPLICANT: DAUBERIES, PIERRE
: TITLE OF INVENTION: MALARIAL PRE-ERYTHROCYTIC STAGE POLYPEPTIDE MOLECULES
: FILE REFERENCE: 0660-0125-0 PCT
:
CURRENT APPLICATION NUMBER: US/08/973.462B
CURRENT FILING DATE: 1998-02-06
:
EARLIER APPLICATION NUMBER: PCT/FR96/00894
EARLIER FILING DATE: 1996-06-12
:
EARLIER APPLICATION NUMBER: FR 95/07007
EARLIER FILING DATE: 1995-06-13
:
NUMBER OF SEQ ID NOS: 29
:
SOFTWARE: PatentIn Ver. 2.0
:
SEQ ID NO: 1
:
LENGTH: 6152
:
TYPE: DNA
:
ORGANISM: P. falciparum
:
US-08-973-462-1

```

	Query Match	6.9%	Pred. 46.4:	DB 4,	Length 6152:
	Best Local Similarity	48.0%:	Pred. NO. 0.42,		
	Matches 169;	Conservative	0;	Mismatches 176;	Indels 7; Gaps 1.
QY	ATATTATTGGTAATAGTGAACCTGTACTACCAAAATAATGACAGCTGGAACCTGCCTTACC	375			
Db	ATATATATATATATATATGATCTTTTACAAAATTTTAAAAATTTTAAATTTTATATATTT	5779			
QY	AATCTTGAAATTTGACCAACAAGTGCTTATATATATGACAGTCATATGTAAATCCAGAAC	435			
Db	AATATTATTATTTTTCCCATATATATATTTTATTTCAATATTTTATTTTAAATTAATGAAG	5839			
QY	TGGACTCGATCGTTAAATTTATTTATGTGTACATTCAAATGTGTGCAATTAATATGCTT	495			
Db	TTTTTTACAGATTTATGTTTTTTTATTAATATATAGATTTCTGTAGAATA-----CT	5892			
QY	CCACAGTAAATATCTGAATAACGATTTTGATGTGAAGCTGCCCTTCTTACTTGGAGT	555			
Db	GTAATATTTTCAATCAGATATATGTAATTTAATTTTGTGTATTATTAATTTATATT	5952			
QY	CTTGATCATACATACCTTTTTATGAGCATGAAATAAACATTTTAACTGATTTCTTA	615			
Db	ATAGAGATTAATAAATTAGCTATTT	6012			
QY	ACTTTGACATTTCAAATTTCTTCTCTTTTCTTTCTTTTCTTTTCTTTTCTTTTCTTTT	667			
Db	TGCTGTATGCAAAATTAAGCTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTCAT	6064			

RESULT 8
US-08-451-405A-2
; Sequence 2, Application US/08451405A

Patent No. 5736358
GENERAL INFORMATION:
APPLICANT: FASEL, NICOLAS JOSEPH
APPLICANT: REYMOND, CHRISTOPHE DOMINIQUE
TITLE OF INVENTION: DICTYOSTELID EXPRESSION VECTOR AND
METHOD OF INVENTION: METHOD FOR EXPRESSING A DESIRED PROTEIN
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: THE WEBB LAM FIRM
STREET: 700 KOPEPS BUILDING, 436 SEVENTH AVENUE
CITY: PITTSBURGH
STATE: PENNSYLVANIA
COUNTRY: UNITED STATES OF AMERICA
ZIP: 15219-1818
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" FLOPPY DISK
COMPUTER: Midwest Micro 486-50
OPERATING SYSTEM: DOS
SOFTWARE: WORDPERFECT 6.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/451,405A
FILING DATE: 26-MAY-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/965,273
FILING DATE: 15-JAN-1993
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 731
TYPE: NUCLEIC ACID
STRANDEDNESS: SINGLE
TOPOLOGY: UNKNOWN
US-08-451-405A-2

Query Match	6.98;	Score 46.2;	DB 1;	Length 731;
Best Local Similarity	44.38;	Pred. No. 0.36;		
Matches 276;	Conservative	0;	Mismatches 343;	Indels 4;
			Gaps	2

QY	9	TTTTCTTTTATTTTAAATTAAGCTTATTAATAGTGGAAAAAATATCTGTGCTCT	68
Db	83	TTATTTTATTTTATTTTAAAAAATTAATAATTAATAATATTTCTATTGAG	142
QY	69	TTAAATTCATTTTATTTTAAATTAAGTATTTTCGAAATGAAACAAGATTGAAAAATTATTTA	128
Db	143	GAGTTTATTTATTTATTTTAAATTTATTTAAATCATTAGAACCTTAAATATGATTTGTGA	202
QY	129	GAAATTTTCTGTGCTTTTCCGTTCATATAGAGAGATGGCTGATGCATPATGGGA	188
Db	203	CGGATATGATTAAGAAAATTTTAAAAAATAATTCAGTAAATTTTGTGATTGAAACAC	266
QY	189	ACTGAATGCATTAAGTATTAATTTTGAAGATTTTGAAGAAAGGAAATGCAATGACAC	248
Db	263	AACCA-----AATTA	318
QY	249	AAATTACAAATGTGTGCGCTGGGACGAAGACATTTGAAGGTCATAGCTTTTACTT	308
Db	320	AAATCAAAAAAAAAAAGTATTTTAAAGAAATTTTAAATTTATATATCTTTAAT	379
QY	309	TAAATCATATATTTGTATATAGTAAACCTGTACTCAAAATATAGCAGCTTGAACTGG	368
Db	380	GTGCAAAACACATTTTAACACACTCTATTTATCTTACAAAGCTTTAAATTTTAAATTTT	433
QY	369	CTTTACCAATCTGAATTTTGACACACAGTGTCTATATATGACAGATCTAATGTAATAATC	428
Db	440	TTTATTTAATATATTTTATTTTAAATTAATTTTATTTTAAATTTTAAATTTTATTTT	495
QY	429	CAGAACTGGACCTCATGCTGAATTAATTTATGTAACCTCAAAATGTCGATTAATAA	488
Db	500	TTTACATCAACCCCTTTAATCAACAATAACATTTTATTTTATTTATTTATATATA	555
QY	489	TATGCTTCACAGTAATAATCTGAATAAACTGATTTGTGATGAAGACGCTCTTCTAATTTA	548
Db	560	T-CATTTGGAATTAATAATATTTTCTCAATGTAGTATATATATTTCTTTTAAATATAA	618


```

; TITLE OF INVENTION: phosphate synthetase II
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon & Vanderhye PC
; STREET: 1100 No. 5849573th Glebe Road, 8th Floor
; City: Arlington
; STATE: Virginia
; COUNTRY: USA
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.24
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/446,855A
; FILING DATE: 06-Jul-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Mitchard, Leonard C
; REGISTRATION NUMBER: 29,009
; REFERENCE/DOCKET NUMBER: 47-80
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-816-4000
; TELEFAX: 703-816-4100
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8920 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: genomic
; US-08-446-855A-1

Query Match          6.5%, Score 43.6; DB 2; Length 8920;
Best Local Similarity 50.0%; Pred. No. 1.6;
Matches 109; Conservative 0; Mismatches 109; Indels 0; Gaps 0;

QY 448 TTAATATTATTATGTGTAACATTCGAATGCTGCATTAATAATGCTCCACAGTAAAT 507
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 834 TAAACTCTTTTAAAGCATTATGATGATGAGAATATCCCAATATATATATATATATA 775
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 508 CTGAACAACGTATTTGGTATTTGAAGCTGCGCTTCTATTTACTTGAGTCTTGACATACA 567
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 774 TTATATATATATATCATATATATTTTCCCATTTTCTTTTTTTATATACATTATA 715
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 568 TACTTTTTATGAGCTATGAAATTAACATTTTAACTGATTTCTTAACTTGCATTT 627
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 714 TTATGTTTAAATATTTATATAATTTACATATACAAAGTTCATTTTCATATGTAATTT 655
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 628 CAATTTCTCTCTCTTTCTTTCTTTCTTTCTTTTCTTTTCTTTT 665
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 654 TTTTCTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 617
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 13
US-09-150-741-1/c
; Sequence 1, Application US/09150741
; Patent No. 6183996
; GENERAL INFORMATION:
; APPLICANT: Stewart et al.
; TITLE OF INVENTION: Nucleotide Sequence Encoding Carbamoyl Phosphate
; Patent No. 6183996
; FILE REFERENCE:
; TITLE OF INVENTION: Synthetase II
; CURRENT APPLICATION NUMBER: US/09/150,741
; EARLIER APPLICATION NUMBER: PL6380
; EARLIER FILING DATE: 1992-12-16
; EARLIER APPLICATION NUMBER: AU93/00617
; EARLIER FILING DATE: 1993-12-02
; EARLIER APPLICATION NUMBER: 08/446,855
; EARLIER FILING DATE: 1995-07-06

```

NUMBER OF SEQ ID NOS: 15
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1
LENGTH: 8920
TYPE: DNA
ORGANISM: Plasmodium falciparum
US-09-150-741-1

Query Match
Best Local Similarity 50.0%; Score 43.6; DB 4; Length 8920;
Pred. No. 1.6;
Matches 109; Conservative 0; Mismatches 109; Indels 0; Gaps 0;

Qy 448 TTTAAATTTTATGCTGATTCGAATGATGCTGATTAATATGCTTCACAGTAAT 507
Db 834 TAAACTGTTTAAAGACTTATGATTCAGAGATATCCCAATATATATATATATA 775
Qy 508 CTGAAACTGATTTGCTGATTCGAAGCTGCTTCTATTTACTGCTGTACATACA 567
Db 774 TTTATATATATATCTATATATATTTTCCCATTTTCTTTTATATACATTATA 715
Qy 568 TACTTTTATAGCTATGAAATTAACATTTTAACTGATTTTCACTTGACATTT 627
Db 714 TTTATGTTAAATATTTTAAATTTACATATACAGTTTCAATATGTAATTTT 655
Qy 628 CAAATTTCTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTT 665
Db 654 TTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTT 617

RESULT 14
US-09-426-290-1
Sequence 1, Application US/09426290
Patent No. 6410712

GENERAL INFORMATION:
APPLICANT: Berglund Ran Oiafsdottir
APPLICANT: Jeffrey Guicher
TITLE OF INVENTION: HUMAN NARCOLEPSY GENE
FILE REFERENCE: 2345.2001-000
CURRENT APPLICATION NUMBER: US/09/426,290
CURRENT FILING DATE: 1999-10-25
NUMBER OF SEQ ID NOS: 24
SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 1
LENGTH: 168575
TYPE: DNA
ORGANISM: Homo Sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (21181)...(21403)
NAME/KEY: CDS
LOCATION: (95252)...(95430)
NAME/KEY: CDS
LOCATION: (101753)...(101996)
NAME/KEY: CDS
LOCATION: (110324)...(110439)
NAME/KEY: CDS
LOCATION: (124058)...(124278)
NAME/KEY: CDS
LOCATION: (127009)...(127130)
NAME/KEY: CDS
LOCATION: (128910)...(129139)
US-09-426-290-1

Query Match
Best Local Similarity 48.0%; Score 43.4; DB 4; Length 168575;
Pred. No. 2.4;
Matches 154; Conservative 0; Mismatches 166; Indels 1; Gaps 1;

Qy 347 AATATAGCAGCTTGAACCTGCTTACCAATCTTGAATTTGACCACAACTGCTTATA 406
Db 120800 AATATATATATATAGAGATCTTTATATCTCTGCAATGTTATCATATTTTCAAGTATA 120859
Qy 407 TATGAGATCTATATGTAATATCCAGACTTGACATCCATCGTTAAATATATATATGCTGA 466

Db 120860 TATATATATATATGAAATATATATATATATATATATATATATATATAT 120919

Qy 467 ACATTCAAATGCTGCTGATTAATATGCTTCACAGTAATATGTAATATGCTGTA 526

Db 120920 TATATATATATATGAAATATATATATATATATATATATATATATATAT 120979

Qy 527 TTGAAAGCTGCTTCTATTTACTGAGTCTGTACATACATCTTTTATGAGCTATG 586

Db 120980 TATATATATATATGAAATATA-TATCTTATATATATATATATATGAAATGATGCTTAA 121038

Qy 587 AATTAACATTTTAAAGTAATTTCTTAACTTGACATTTCAATTTCTCTCTTTT 646

Db 121039 AACACATTTCTTTGATTTGAAACTTTTAAATTTTATTTTCCATTTATTTGCTAGTATGT 121098

Qy 647 CTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTT 667

Db 121099 AGAAGTATATGATTTTCTTT 121119

RESULT 15
US-08-487-826B-13/C

Sequence 13, Application US/08487826B
Patent No. 5993827

GENERAL INFORMATION:

APPLICANT: Sim, Kim L.
APPLICANT: Chitans, Chetan
APPLICANT: Miller, Louis H.
APPLICANT: Peterson, David S.
APPLICANT: Su, Xin-zhaun
APPLICANT: Wellens, Thomas E.
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
NUMBER OF SEQUENCES: 45
CORRESPONDENCE ADDRESSES:
ADDRESS: Knobb Martens Olson & Bear
STREET: 620 Newport Center Drive 16th Floor
CITY: Newport Beach
STATE: California
COUNTRY: US

ZIP: 92660
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/487,826B
FILING DATE: 10-SEP-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Israelson, Ned

REGISTRATION NUMBER: 29,655
REFERENCE/DOCKET NUMBER: NIH121.001CPI
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176

INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 19124 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear

MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-487-826B-13

Query Match
Best Local Similarity 43.9%; Score 43; DB 2; Length 19124;
Pred. No. 2.2;
Matches 293; Conservative 0; Mismatches 360; Indels 14; Gaps 2;

Qy 2 AATATACCTTTTCTTTTATTAATTAAGCTTATATAATGTCGAAATAATATATCTTG 61

THIS PAGE BLANK (USPTO)

Accession	Sequence	Position
Db	1 AATATACTTTTCTTTTATTTAATAAGCTTATAATCGGAGAAAAATATCTT	60
QY	61 GTGTTCCTTAATTCATTTTATTATACTATTTTCAGATGAAATAAAGATTGAAA	120
Db	61 GTGTTCCTTAATTCATTTTATTTAATACTATTTTCAGATGAAATAAAGATTGAAA	120
QY	121 ATATATTAGAAATTTTTCGTGCTTTTTCCTGTTCAGATTAAGAGAGGAGATG	180
Db	121 ATATATTAGAAATTTTTCGTGCTTTTTCCTGTTCAGATTAAGAGAGATG	180
QY	181 CATAGGGAAGCTCAATGCATACTATATATTGAAGTTATGAAAGAGGAAATACAA	240
Db	181 CATAGGGAAGCTCAATGCATACTATATATTGAAGTTATGAAAGAGGAAATACAA	240
QY	241 ATGACACAAATTCAAATGTGTGTGTGGGAGGAACATCTTTTAAGGTCATGAGTT	300
Db	241 ATGACACAAATTCAAATGTGTGTGTGGGAGGAACATCTTTTAAGGTCATGAGTT	300

Db 241 ATGACACAAATTTACAAATGTGTGCGTGGGACGAACATCTTTGAAGTCATGAGTT 300
Qy 301 TGTAGTTAAACATCATATATTTGTAATGTAACCTGTACTCAAAATATATAAGCAGTT 360
Db 301 TGTAGTTAAACATCATATATTTGTAATGTAACCTGTACTCAAAATATATAAGCAGTT 360
Qy 361 GAACTGGCTTTACCAATCTTGAATTTTGACCAAGTCTTATATATGACAGATCTAAT 420
Db 361 GAACTGGCTTTACCAATCTTGAATTTTGACCAAGTCTTATATATGACAGATCTAAT 420
Qy 421 GTAAATCCAGAACTGGACCTCATCGTTAAATTTATATGTATGTAACATTCAAATGTGT 480
Db 421 GTAAATCCAGAACTGGACCTCATCGTTAAATTTATGTATGTAACATTCAAATGTGT 480
Qy 481 GCATTAAATATGCTTCACAGTAATAATCTGAAAACTGATTTGTGATTAAGCTGCTT 540
Db 481 GCATTAAATATGCTTCACAGTAATAATCTGAAAACTGATTTGTGATTAAGCTGCTT 540
Qy 541 TCTATTACTGAGCTTGTGTACATACATACCTTTTATGAGCTANGAATAAACAATTTT 600
Db 541 TCTATTACTGAGCTTGTGTACATACATACCTTTTATGAGCTANGAATAAACAATTTT 600
Qy 601 AAACGAAATTTCTTACCTTGACATTTCAAAATTTCTTCTTTCTTTCTTTCTTTT 660
Db 601 AAACGAAATTTCTTACCTTGACATTTCAAAATTTCTTCTTTCTTTCTTTCTTTT 660
Qy 661 TTTTGTGAGA 672
Db 661 TTTTGTGAGA 672

RESULT 2

US-09-880-107-2421
; Sequence 2421, Application US/09880107
; Patent No. US20020142981A1
; GENERAL INFORMATION:
; APPLICANT: Horne, Darci T.
; APPLICANT: Vockley, Joseph G.
; APPLICANT: Scherf, Uwe
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Gene Expression Profiles In Liver Cancer
; FILE REFERENCE: 44921-5028-WO
; CURRENT APPLICATION NUMBER: US/09/880,107
; PRIOR FILING DATE: 2001-06-14
; PRIOR APPLICATION NUMBER: US 60/211,379
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: US 60/237,054
; NUMBER OF SEQ ID NOS: 3950
; SOFTWARE: Patent Ver. 2.1
; SEQ ID NO 2421
; LENGTH: 672
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20020142981A1 M93036
US-09-880-107-2421

Query Match 100.0%; Score 672; DB 10; Length 672;
Best Local Similarity 100.0%; Pred. No. 5e-105;
Matches 672; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AATATACTTTTCTTTTATTTTAAATAGTCTATTAATGCGGAAAAATATATCTT 60
Db 1 AATATACTTTTCTTTTATTTTAAATAGTCTATTAATGCGGAAAAATATATCTT 60
Qy 61 GGTGTCCTTTAATTCATTTTATTAATAGTCTATTAATGCGGAAAAATATATCTT 120
Db 61 GGTGTCCTTTAATTCATTTTATTAATAGTCTATTAATGCGGAAAAATATATCTT 120
Qy 121 ATTATTTAATTTTCTGCTTTTCCGCTTTTCAATATAAGAGATGGTGAGTG 180
Db 121 ATTATTTAATTTTCTGCTTTTCCGCTTTTCAATATAAGAGATGGTGAGTG 180

Qy 181 CATAGGAACTCAATCATATATTTTGAAGATTATAGAAAGGAAAAATAGCAA 240
Db 181 CATAGGAACTCAATCATATATTTTGAAGATTATAGAAAGGAAAAATAGCAA 240
Qy 241 ATGACACAAATTTACAAATGTGTGCGTGGGACGAACATCTTTGAAGTCATGAGTT 300
Db 241 ATGACACAAATTTACAAATGTGTGCGTGGGACGAACATCTTTGAAGTCATGAGTT 300
Qy 301 TGTAGTTAAACATCATATATTTGTAATGTAACCTGTACTCAAAATATATAAGCAGTT 360
Db 301 TGTAGTTAAACATCATATATTTGTAATGTAACCTGTACTCAAAATATATAAGCAGTT 360
Qy 361 GAACTGGCTTTACCAATCTTGAATTTTGACCAAGTCTTATATATGACAGATCTAAT 420
Db 361 GAACTGGCTTTACCAATCTTGAATTTTGACCAAGTCTTATATATGACAGATCTAAT 420
Qy 421 GTAAATCCAGAACTGGACCTCATCGTTAAATTTATATGTATGTAACATTCAAATGTGT 480
Db 421 GTAAATCCAGAACTGGACCTCATCGTTAAATTTATGTATGTAACATTCAAATGTGT 480
Qy 481 GCATTAAATATGCTTCACAGTAATAATCTGAAAACTGATTTGTGATTAAGCTGCTT 540
Db 481 GCATTAAATATGCTTCACAGTAATAATCTGAAAACTGATTTGTGATTAAGCTGCTT 540
Qy 541 TCTATTACTGAGCTTGTGTACATACATACCTTTTATGAGCTANGAATAAACAATTTT 600
Db 541 TCTATTACTGAGCTTGTGTACATACATACCTTTTATGAGCTANGAATAAACAATTTT 600
Qy 601 AAACGAAATTTCTTACCTTGACATTTCAAAATTTCTTCTTTCTTTCTTTCTTTT 660
Db 601 AAACGAAATTTCTTACCTTGACATTTCAAAATTTCTTCTTTCTTTCTTTCTTTT 660
Qy 661 TTTTGTGAGA 672
Db 661 TTTTGTGAGA 672

RESULT 3

US-09-777-564-191/C
; Sequence 191, Application US/09777564
; Patent No. US2002022591A1
; GENERAL INFORMATION:
; APPLICANT: Algate, Paul A.
; APPLICANT: Mannion, Jane
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF OVARIAN CANCER
; FILE REFERENCE: 210121.493
; CURRENT APPLICATION NUMBER: US/09/777,564
; PRIOR FILING DATE: 2001-02-05
; NUMBER OF SEQ ID NOS: 1730
; SOFTWARE: FastSeq for Window Version 4.0
; SEQ ID NO 191
; LENGTH: 621
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)-(621)
; OTHER INFORMATION: n = A,T,C or G
US-09-777-564-191

Query Match 68.7%; Score 461.6; DB 10; Length 621;
Best Local Similarity 91.3%; Pred. No. 1.1e-69;
Matches 525; Conservative 0; Mismatches 44; Indels 6; Gaps 4;

Qy 26 AATAAGCTGTATTAATGCGGAAAAATATCTGTCTCTTAATTTCAATTTTAT 85
Db 569 AATAAGTGTATTAATGCGGAAAAATTTCTGTG---TTCNTTAATTTCAATTTTAT 513
Qy 86 TAATATATTTTCAAGATGACAAAAGATGAAAAATTTATTAAGATTTTCTGCTGCT 145
Db 512 TAANGCTATTTTCAAGATG-ACNAAGATTTGAAAAATTTATTAAG-ATTTTCTGCTGCT 455

[illegible]

Query Match	Similarity	Score	DB 10:	length
Best Local	Similarity	99.88:	Pred. No. 1.4e-67:	
Matches	450:	Conservative	0:	Mismatches 1: indels 0: Gaps 0:
QY	156	TCAGATAAAGAGATGGTGAGATGCATATAGGAACTCAATGCATACATATATATATTTGAA	215	
Db	1078	TCAGATAAAGAGATGGTGAGATGCATATAGGAACTCAATGCATACATATATATATTTGAA	1137	
QY	216	GATTATAGAGAAGGGGAAATAGCAAAATGACACAAATTCAAATGCTGTGGTGGGACG	275	
Db	1138	GATTATAGAGAAGGGGAAATAGCAAAATGACACAAATTCAAATGCTGTGGTGGGACG	1197	
QY	276	AAGACATCTTTGAAGTCATGATGTTGTTAGTTAACTCATCATATATTTTGAATAGGAAA	335	
Db	1198	AAGACATCTTTGAAGTCATGATGTTGTTAGTTAACTCATCATATATTTTGAATAGGAAA	1257	
QY	336	CGTGATCTCAAAATATATAGACGCTTGAACCTGCTTACCAATCTTGAATTTGACCACA	395	
Db	1258	CGTGATCTCAAAATATATAGACGCTTGAACCTGCTTACCAATCTTGAATTTGACCACA	1317	
QY	396	AGTGCTTATATATGACAGATCTAATGTAAATTCAGAACTGGACATCGTTAAATT	455	

Db 1318 AGTGTCTATATATGACAGATCTAATGTAAATTCAGAACTTGACCTCATCTTAAAT 1377
Qy 456 ATTATAGTGAATCAATGATGTCATTAATATGCTTCACAGTAAATCTGAAAA 515
Db 1378 ATTATAGTGAATCAATGATGTCATTAATATGCTTCACAGTAAATCTGAAAA 1437
Qy 516 CTGATTTGTGATGAAAGCTGCTTCTATTTACTTGAGTCTTGTACATCACTTTT 575
Db 1438 CTGATTTGTGATGAAAGCTGCTTCTATTTACTTGAGTCTTGTACATCACTTTT 1497
Qy 576 TATGAGCTATGAATAAATCAATTTAACTG 606
Db 1498 TATGAGCTATGAATAAATCAATTTAACTG 1528

RESULT 6

US-09-920-300A-14/C
; Sequence 14, Application US/09920300A
; Patent No. US20020136728A1
; GENERAL INFORMATION:
; APPLICANT: King, Gordon E.
; APPLICANT: Meagher, Madeleine Joy
; APPLICANT: Xu, Jiangchun
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.547
; CURRENT APPLICATION NUMBER: US/09/920.300A
; NUMBER OF SEQ ID NOS: 1789
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14
; LENGTH: 637
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-920-300A-14

Query Match 66.0%; Score 443.4; DB 10; Length 637;
Best Local Similarity 99.8%; Pred. No. 1.3e-66;
Matches 444; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 156 TCAGATAAGAGATGGGTGATGATGATAGGAACTCATGATATATTAATTTGAA 215
Db 445 TGAGATTAAGAGATGGGTGATGATGATAGGAACTCATGATATATTAATTTGAA 386
Qy 216 GATTATAGAAGAGGAAATAGCAAAATGACACAATTAACAATGTGTGCGGGGAGC 275
Db 385 GATTATAGAAGAGGAAATAGCAAAATGACACAATTAACAATGTGTGCGGGGAGC 326
Qy 276 AAGACATCTTTGAAGTCAATGATGATTTGATTAACTCATATATTTGTAATAGTAAA 335
Db 325 AAGACATCTTTGAAGTCAATGATGATTTGATTAACTCATATATTTGTAATAGTAAA 266
Qy 336 CCTGACTCAAAATATTAAGAGCTTGAAGCTTTTACCAATCTGTAATTTGACACA 395
Db 265 CCTGACTCAAAATATTAAGAGCTTGAAGCTTTTACCAATCTTGAATTTGACACA 206
Qy 396 AGTGTCTATATATGACAGATCTAATGTAAATCCAGAACTGAGCTCATGCTTAAAT 455
Db 205 AGTGTCTATATATGACAGATCTAATGTAAATCCAGAACTGAGCTCATGCTTAAAT 146
Qy 456 ATTATATGTAACATTTCAATGTGTCATTAATATGCTTCCACAGTAAATCTGAAAA 515
Db 145 ATTATATGTAACATTTCAATGTGTCATTAATATGCTTCCACAGTAAATCTGAAAA 86
Qy 516 CTGATTTGTGATGAAAGCTGCTTCTATTTACTTGAGTCTTGTACATCACTTTT 575
Db 85 CTGATTTGTGATGAAAGCTGCTTCTATTTACTTGAGTCTTGTACATCACTTTT 26
Qy 576 TATGAGCTATGAATAAATCAATTT 600
Db 25 TATGAGCTATGAATAAATCAATTT 1

RESULT 7

US-10-033-528-14/C
; Sequence 14, Application US/10033528
; Patent No. US20020131971A1
; GENERAL INFORMATION:
; APPLICANT: King, Gordon E.
; APPLICANT: Meagher, Madeleine Joy
; APPLICANT: Xu, Jiangchun
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.547C1
; CURRENT APPLICATION NUMBER: US/10/033.528
; NUMBER OF SEQ ID NOS: 1896
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14
; LENGTH: 637
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-033-528-14

Query Match 66.0%; Score 443.4; DB 12; Length 637;
Best Local Similarity 99.8%; Pred. No. 1.3e-66;
Matches 444; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 156 TCAGATAAGAGATGGGTGATGATGATAGGAACTCATGATATATTAATTTGAA 215
Db 445 TGAGATTAAGAGATGGGTGATGATGATAGGAACTCATGATATATTAATTTGAA 386
Qy 216 GATTATAGAAGAGGAAATAGCAAAATGACACAATTAACAATGTGTGCGGGGAGC 275
Db 385 GATTATAGAAGAGGAAATAGCAAAATGACACAATTAACAATGTGTGCGGGGAGC 326
Qy 276 AAGACATCTTTGAAGTCAATGATGATTTGATTAACTCATATATTTGTAATAGTAAA 335
Db 325 AAGACATCTTTGAAGTCAATGATGATTTGATTAACTCATATATTTGTAATAGTAAA 266
Qy 336 CCTGACTCAAAATATTAAGAGCTTGAAGCTTTTACCAATCTTGAATTTGACACA 395
Db 265 CCTGACTCAAAATATTAAGAGCTTGAAGCTTTTACCAATCTTGAATTTGACACA 206
Qy 396 AGTGTCTATATATGACAGATCTAATGTAAATCCAGAACTGAGCTCATGCTTAAAT 455
Db 205 AGTGTCTATATATGACAGATCTAATGTAAATCCAGAACTGAGCTCATGCTTAAAT 146
Qy 456 ATTATATGTAACATTTCAATGTGTCATTAATATGCTTCCACAGTAAATCTGAAAA 515
Db 145 ATTATATGTAACATTTCAATGTGTCATTAATATGCTTCCACAGTAAATCTGAAAA 86
Qy 516 CTGATTTGTGATGAAAGCTGCTTCTATTTACTTGAGTCTTGTACATCACTTTT 575
Db 85 CTGATTTGTGATGAAAGCTGCTTCTATTTACTTGAGTCTTGTACATCACTTTT 26
Qy 576 TATGAGCTATGAATAAATCAATTT 600
Db 25 TATGAGCTATGAATAAATCAATTT 1

RESULT 8

US-09-920-300A-1559/C
; Sequence 1559, Application US/09920300A
; Patent No. US20020136728A1
; GENERAL INFORMATION:
; APPLICANT: King, Gordon E.
; APPLICANT: Meagher, Madeleine Joy
; APPLICANT: Xu, Jiangchun
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.547
; CURRENT APPLICATION NUMBER: US/09/920.300A

OY	216	GATTATTAGAAAGGGAATATGCAATTCGACACAATTTACAAATGTGTGGCGCGGGACG	275
Db	391	CATTATTAGAAAGGGAATATGCAATTTGACACAATTTACAAATGTGTGGCGGGACN	332
OY	276	AAGACATCTTTGAAGGTCATGAGTTTGTATTGTTTACATCATATATTTGTAATAGTGA	335
Db	331	AAGACATCTTTGAAGGTCATGAGTTTGTATTGTTTACATCATATATTTGTAATAGTGA	277
OY	336	CCTGTACCTCAAAATATAGCAGCTTGAACCTGCTTACCAATCTTAAATTTGACACA	395
Db	271	CCTGTACCTCAAAATATAGCAGCTTGAACCTGCTTACCAATCTTGAATTTGACACA	212
OY	396	AGTGCCTATATATGACAGATCTATATGTAATTCAGAACTTGGACCTCCATCGTTAAAT	455
Db	211	AGTGCCTATATATGACAGATCTATATGTAATTCAGAACTTGGACCTCCATCGTTAAAT	152
OY	456	ATTATATGTAACATTCCAATGTGTGCAATTAATATGCTTCCACAGTAATCTGAAAA	515
Db	151	ATTATATGTAACATTCCAATGTGTGCAATTAATATGCTTCCACAGTAATCTGAAAA	92
OY	516	CTGATTTGTGATTGAAGCTGCTTTCTATATTTACTTGATGCTTGTACATCACTATTT	575
Db	91	CTGATTTGTGATTGAAGCTGCTTTCTATATTTACTTGATGCTTGTACATCACTATTT	32
OY	576	TATGAGCATGAATTAACATTTT	601
Db	31	TATGAGCATGAATTAACATTTT	6

```

RESULT 11
US-09-777-564-1254/C
: Sequence 1254, Application US/09777564
: Patent No. US20020022591A1
: GENERAL INFORMATION:
: APPLICANT: Algateo, Paul A.
: APPLICANT: Mannito, Jane
: TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
: TITLE OF INVENTION: AND DIAGNOSIS OF OVARIAN CANCER
: FILE REFERENCE: 210121.493
: CURRENT APPLICATION NUMBER: US/09/777,564
: CURRENT FILING DATE: 2001-02-05
: NUMBER OF SEQ ID NOS: 1730
: SOFTWARE: FastSeq for Window Version 4.0
: SEQ ID NO 1254
: LENGTH: 439
: TYPE: DNA
: ORGANISM: Homo sapiens
: US-09-777-564-1254

```

Query Match	65.3%;	Score 439;	DB 10;	Length 439;
Best Local Similarity	100.0%;	Pred. No. 6.6e-66;		
Matches 439;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

OY	162	AAAGAGATGGGTGAGATGCACTAGGGAACCTCAATGCAATCTATATATATTTGAAGATTAT	221
Dp	439	AAAGAGATGGGTGAGATGCACTAGGGAACCTCAATGCAATCTATATATTTGAAGATTAT	380
OY	222	AGAAAGAGGGAATAGCAAAATGACACAAATTACAAATGTGTGTCCTGGGAGCAAGACA	281
Dp	379	AGAAAGAGGGAATAGCAAAATGACACAAATTACAAATGTGTGTCCTGGGAGCAAGACA	320
OY	282	TCTTTGAAGGATGAGCTTGTGTGTTAAATCATATATTTGTAATAGGAACCTGTA	341
Dp	319	TCTTTGAAGGATGAGCTTGTGTGTTAAATCATATATTTGTAATAGGAACCTGTA	266
OY	342	CTCAAAATATAGCAGCTTGAACCTGGCTTTACCAATCTTGAATTTGACCACAAAGTGC	401
Dp	259	CTCAAAATATAGCAGCTTGAACCTGGCTTTACCAATCTTGAATTTGACCACAAAGTGC	200
OY	402	TTTATATGCGATCTATATGTAATAATCCAGACCTGGACCCATCGTTAAATTTATTTAT	461
Dp	199	TTTATATGCGATCTATATGTAATAATCCAGACCTGGACCCATCGTTAAATTTATTTAT	140

QY	462	GGTGAACCTTCAAAATGTCGATTAAATATGCTTCACAGTAAATCTGAAAAATCATTT	521
Db	139	GGTGAACCTTCAAAATGTCGATTAAATATGCTTCACAGTAAATCTGAAAAATCATTT	80
QY	522	TGTGATTTGAAGCGCCCTTCTATTACTGAGCTGATCATATCACTTTTATGAG	581
Db	79	TGTGATTTGAAGCGCCCTTCTATTACTGAGCTGATCATATCACTTTTATGAG	20
QY	582	CTATGAATATAACATTTT	600
Db	19	CTATGAATATAACATTTT	1

```

RESULT 12
US-09-998-598-290/c
: Sequence 290. Application US/09998598
: Patent No. US20020150922A1
GENERAL INFORMATION:
: APPLICANT: Stolk, John A.
: APPLICANT: Xu, Jiangchun
: APPLICANT: Chesnault, Ruth A.
: APPLICANT: Meagher, Madeline Joy
: TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
: TITLE OF INVENTION: DIAGNOSIS OF COLON CANCER
: FILE REFERENCE: 210121.561
: CURRENT APPLICATION NUMBER: US/09/998,598
: CURRENT FILING DATE: 2001-11-16
: NUMBER OF SEQ ID NOS: 2606
: SOFTWARE: Corlax Invention Disclosure Database
: SEQ ID NO 290
: LENGTH: 534
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: 1, 403
: OTHER INFORMATION: n = A,T,C or G
: US-09-998-598-290

```

Query Match	64.58;	Score 433.4;	DB 10;	length 534;
Best Local Similarity	99.68;	Pred. NO. 5.9e-65;		
Matches 445;	Conservative 0;	Mismatches 1;	Indels 1;	Gaps 1;

Qy	156	TCACATTAAGAGATGGGTGGATGCAATAGGAACCTCAATGCAATAC- TATATATTTGA	21.4
Db	450	TGACATTAAGAGATGGGTGGATGCAATAGGAACCTCAATGCAATAC- TATATATTTGA	39.1
Qy	215	AGATTATAGAAAGGGGAATAGCAAAATGACACAAATTCGATGCTGGCGAC	27.4
Db	390	AGATTATAGAAAGGGGAATAGCAAAATGACACAAATTCGATGCTGGCGAC	33.1
Qy	275	GAACACATCTTTGAAGGTCATGAGTTTGTATAGTTAACTATCATATATTGTAATGTGA	33.4
Db	330	GAACACATCTTTGAAGGTCATGAGTTTGTATAGTTAACTATCATATATTGTAATGTGA	27.1
Qy	335	ACCGTACTCAAAATATATACAGCTTGAACCTGGCTTACCAATCTTGAATTTGACAC	39.4
Db	270	ACCGTACTCAAAATATATACAGCTTGAACCTGGCTTACCAATCTTGAATTTGACAC	21.1
Qy	395	AAGTGTCTTATATATGAGATCTTATATGTAATAATCCAGAACTGGACATCGCTTAAAT	45.4
Db	210	AAGTGTCTTATATATGAGATCTTATATGTAATAATCCAGAACTGGACATCGCTTAAAT	15.1
Qy	455	TATTTATGCTAACTCAATATGTCGATTTAAATATAGCTTCCACATTAATCTGAAA	51.4
Db	150	TATTTATGCTAACTCAATATGTCGATTTAAATATAGCTTCCACATTAATCTGAAA	9.1
Qy	515	ACTGATTTGATTTGAAGCTGCCCTTCTATTAAGTGAAGCTCTTGAACATACACTTTT	57.4
Db	90	ACTGATTTGATTTGAAGCTGCCCTTCTATTAAGTGAAGCTCTTGAACATACACTTTT	3.1
Qy	575	TTATGAGCTATGAATAAACAATTTTA	60.1

Db 30 TTATGAGCTATGAATAAATCATTTTA 4

RESULT 13
US-09-998-598-1423

; Sequence 1423, Application US/09998598
; Patent No. US20020150922A1

; GENERAL INFORMATION:

; APPLICANT: Stolk, John A.

; APPLICANT: Xu, Jiangchun

; APPLICANT: Chenault, Ruth A.

; APPLICANT: Meagher, Madeleine Joy

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND

; FILE REFERENCE: 210121.561

; CURRENT APPLICATION NUMBER: US/09/998,598

; NUMBER OF SEQ ID NOS: 2606

; SOFTWARE: Corixa Invention Disclosure Database

; SEQ ID NO 1423

; LENGTH: 566

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: misc_feature

; LOCATION: 565

; OTHER INFORMATION: n = A,T,C or G

; US-09-998-598-1423

Query Match Best Local Similarity 99.1%; Score 430.2; DB 10; Length 566;

Matches 432; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 156 TCAGATAAAGGAGATGGTGAGATGCATAGGCACTCAATGCATATATATTAATTGAA 215
Db 131 TGAGATTAAGGAGATGGTGAGATGCATAGGCACTCAATGCATATATATTAATTGAA 190
QY 216 GATTATAGAGAAGGAATAGCAATAGCACAAATTCGAATGCTGGTGGGAGC 275
Db 191 GATTATAGAGAAGGAATAGCAATAGCACAAATTCGAATGCTGGTGGGAGC 250
QY 276 AAGACATCTTTGAAGCTCATGAGTTGTTAGTTTAACATCATATATTTGTAATAGTAAA 335
Db 251 AAGACATCTTTGAAGCTCATGAGTTGTTAGTTTAACATCATATATTTGTAATAGTAAA 310
QY 336 CCTGACTCAAAATATATAGCAGCTTGAACCTGCTTACCAATCTTGAATTTGACCACA 395
Db 311 CCTGACTCAAAATATATAGCAGCTTGAACCTGCTTACCAATCTTGAATTTGACCACA 370
QY 396 AATGCTTATATATGAGATCTATATGTAATAATCCAACTGGACCTCCATCGTTAAAT 455
Db 371 AATGCTTATATATGAGATCTATATGTAATAATCCAACTGGACCTCCATCGTTAAAT 430
QY 456 ATTATGCTAATCATCAATGTGTCATTAAATATGCTCCACAGTAATAATCTGAAAA 515
Db 431 ATTATGCTAATCATCAATGTGTCATTAAATATGCTCCACAGTAATAATCTGAAAA 490
QY 516 CTGATTTGATTTGAAGAGCTGCTTTCTATTTACTTGATGCTTTGATACATACACTTTT 575
Db 491 CTGATTTGATTTGAAGAGCTGCTTTCTATTTACTTGATGCTTTGATACATACACTTTT 550
QY 576 TATGAGCTATGAATA 591
Db 551 TATGAGCTATGAATA 566

RESULT 14

US-09-922-217-591/c

; Sequence 591, Application US/09922217

; Patent No. US2002076414A1

; GENERAL INFORMATION:

; APPLICANT: Xu, Jiangchun

; APPLICANT: Lodes, Michael J.
; APPLICANT: Secretist, Heather
; APPLICANT: Benson, Darin R.
; APPLICANT: Meagher, Madeleine Joy
; APPLICANT: Stolk, John A.
; APPLICANT: Wang, Tonglong
; APPLICANT: Jiang, Yugu
; APPLICANT: Smith, Carole Lynn
; APPLICANT: King, Gordon E.
; APPLICANT: Wang, Aijun

; APPLICANT: Clapper, Jonathan D.

; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS

; FILE REFERENCE: 210121.471C13

; CURRENT APPLICATION NUMBER: US/09/922,217

; NUMBER OF SEQ ID NOS: 1124

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 591

; LENGTH: 425

; TYPE: DNA

; ORGANISM: Homo sapiens

; US-09-922-217-591

Query Match Best Local Similarity 99.8%; Score 415.4; DB 10; Length 425;

Matches 416; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 156 TCAGATAAAGGAGATGGTGAGATGCATAGGCACTCAATGCATATATTAATTGAA 215
Db 417 TCAGATAAAGGAGATGGTGAGATGCATAGGCACTCAATGCATATATTAATTGAA 358
QY 216 GATTATAGAGAAGGAATAGCAATAGCACAAATTCGAATGCTGGTGGGAGC 275
Db 357 GATTATAGAGAAGGAATAGCAATAGCACAAATTCGAATGCTGGTGGGAGC 298
QY 276 AAGACATCTTTGAAGCTCATGAGTTGTTAGTTTAACATCATATATTTGTAATAGTAAA 335
Db 297 AAGACATCTTTGAAGCTCATGAGTTGTTAGTTTAACATCATATATTTGTAATAGTAAA 238
QY 336 CCTGACTCAAAATATATAGCAGCTTGAACCTGCTTACCAATCTTGAATTTGACCACA 395
Db 237 CCTGACTCAAAATATATAGCAGCTTGAACCTGCTTACCAATCTTGAATTTGACCACA 178
QY 396 AATGCTTATATATGAGATCTATATGTAATAATCCAACTGGACCTCCATCGTTAAAT 455
Db 177 AATGCTTATATATGAGATCTATATGTAATAATCCAACTGGACCTCCATCGTTAAAT 118
QY 456 ATTATGCTAATCATCAATGTGTCATTAAATATGCTCCACAGTAATAATCTGAAAA 515
Db 117 ATTATGCTAATCATCAATGTGTCATTAAATATGCTCCACAGTAATAATCTGAAAA 58
QY 516 CTGATTTGATTTGAAGAGCTGCTTTCTATTTACTTGATGCTTTGATACATACACTTT 572
Db 57 CTGATTTGATTTGAAGAGCTGCTTTCTATTTACTTGATGCTTTGATACATACACTTT 1

RESULT 15

US-09-833-263-591/c

; Sequence 591, Application US/09833263

; Patent No. US20020110547A1

; GENERAL INFORMATION:

; APPLICANT: Wang, Aijun

; APPLICANT: Clapper, Jonathan D.

; APPLICANT: Stolk, John A.

; APPLICANT: Meagher, Madeleine J.

; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND

; FILE REFERENCE: 210121.471C12

; CURRENT APPLICATION NUMBER: US/09/833,263

; NUMBER OF SEQ ID NOS: 1093

; SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 591
LENGTH: 425
TYPE: DNA
ORGANISM: Homo sapien
US-09-833-263-591

Query Match 61.8%; Score 415.4; DB 10; Length 425;
Best Local Similarity 99.8%; Pred. No. 6e-62;
Matches 416; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY	156	TCAGATAAAGGAGTGGGTGAGATGCATAGGAACTCAATGCATATATATTTGAA	215
DB	417	TCAGATAAAGGAGTGGGTGAGATGCATAGGAACTCAATGCATATATATTTGAA	358
QY	216	GATTATAGAGAAGGAAATAGCAAAATGACACAAATTTACAAATGTGTGCGTGGAGC	275
DB	357	GATTATAGAGAAGGAAATAGCAAAATGACACAAATTTACAAATGTGTGCGTGGAGC	298
QY	276	AAGACATCTTTGAAGTGCATGAGTTGTTAGTTAACATCATATATTTGTAATAGTGAAA	335
DB	297	AAGACATCTTTGAAGTGCATGAGTTGTTAGTTAACATCATATATTTGTAATAGTGAAA	238
QY	336	CCTGTACTCAAAATATPAGCAGCTTGAAGCTTGCAATCTTGCAATCTTGCAATCTTGCA	395
DB	237	CCTGTACTCAAAATATPAGCAGCTTGAAGCTTGCAATCTTGCAATCTTGCAATCTTGCA	178
QY	396	AGTGTCTTAATATATGAGATCTAATGTAAATCCAGAACTTGGACCTGCATCGTTAAATTT	455
DB	177	AGTGTCTTAATATATGAGATCTAATGTAAATCCAGAACTTGGACCTGCATCGTTAAATTT	118
QY	456	ATTATGTGTACATTCAAATGTGTGATTAATATGCTTCCACAGTAATAATCTGAAAAA	515
DB	117	ATTATGTGTACATTCAAATGTGTGATTAATATGCTTCCACAGTAATAATCTGAAAAA	58
QY	516	CTGATTTGTATGAAAGCTGCTTCTATTTACTGAGTCTGTACATACATACCTT	572
DB	57	CTGATTTGTATGAAAGCTGCTTCTATTTACTGAGTCTGTACATACATACCTT	1

Search completed: November 26, 2002, 23:47:34
Job time: 28.289 secs